

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 11:08:05 ; Search time 83.16 Seconds  
(without alignments)  
5685.367 Million cell updates/sec

Title: US-09-800-198-8  
Perfect score: 14581  
Sequence: 1 MDVKDRRHSLTRGCKEC.....ELADSSNIQFLRQNMKGK 2733

Scoring table: BLASUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL19.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriaph.\*
- 17: sp\_archaea.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14316.5	98.2	2764	11 Q9WTS5	Q9WTS5 mus musculus
2	14291	98.0	2765	11 Q9RLK2	Q9RLK2 rattus norv
3	13934.5	95.6	2802	13 Q9DER5	Q9DER5 gallus gall
4	10309	70.7	2715	11 Q9WTS6	Q9WTS6 mus musculus
5	9522	65.3	2825	11 Q70465	Q70465 mus musculus
6	9518	65.3	2771	11 Q9WTS7	Q9WTS7 mus musculus
7	9424	64.6	2590	13 Q9W7R4	Q9W7R4 brachydanio
8	9241.5	63.4	2824	13 Q9W7R3	Q9W7R3 brachydanio
9	9209	63.2	2346	11 Q9JLC1	Q9JLC1 mus musculus
10	9055	62.1	1737	4 Q9ULU2	Q9ULU2 homo sapien
11	8541.5	58.6	2705	13 Q9W6V6	Q9W6V6 gallus gall
12	8484.5	58.2	2731	11 Q9WTS4	Q9WTS4 mus musculus
13	8406.5	57.7	2725	4 Q9UKZ4	Q9UKZ4 homo sapien
14	6822	46.8	1769	4 Q9P273	Q9P273 homo sapien
15	4848	33.2	930	11 Q9JLC0	Q9JLC0 mus musculus
16	4465	30.6	849	4 Q9NT68	Q9NT68 homo sapien

17	4393.5	30.1	1364	4	075981	075981 homo sapien
18	4060.5	27.8	831	13	Q9PU49	Q9PU49 gallus gall
19	3939.5	27.0	2731	5	061307	061307 drosophila
20	3933.5	27.0	2711	5	018366	018366 drosophila
21	3916.5	26.9	2515	5	024551	024551 drosophila
22	3915.5	26.9	2515	5	Q9VNU6	Q9VNU6 drosophila
23	3880.5	26.6	1045	4	Q9NVW1	Q9NVW1 homo sapien
24	3628.5	24.9	964	4	Q9NV77	Q9NV77 homo sapien
25	2661.5	18.3	730	4	Q96MS6	Q96MS6 homo sapien
26	2335	16.0	625	4	Q96SV2	Q96SV2 homo sapien
27	2212	15.2	2560	5	Q21980	Q21980 caenorhabdi
28	2090	14.3	609	11	Q923P3	Q923P3 mus musculu
29	2016	13.8	337	11	Q9RIK0	Q9RIK0 rattus norv
30	1582.5	10.9	272	11	Q9RIJ9	Q9RIJ9 rattus norv
31	1537.5	10.5	266	11	Q9RIK1	Q9RIK1 rattus norv
32	1384	9.5	229	11	Q9QVZ1	Q9QVZ1 mus musculu
33	1328	9.1	442	4	Q9NZJ2	Q9NZJ2 homo sapien
34	1203.5	8.3	560	4	P98202	P98202 homo sapien
35	1174	8.1	777	5	Q24550	Q24550 drosophila
36	1167	8.0	777	5	Q9VYP1	Q9VYP1 drosophila
37	1132	7.8	375	4	Q9P2P4	Q9P2P4 homo sapien
38	1110	7.6	587	5	Q9VYN8	Q9VYN8 drosophila
39	1043.5	7.2	1124	5	Q9VYN6	Q9VYN6 drosophila
40	752.5	5.2	278	13	Q9DEQ8	Q9DEQ8 gallus gall
41	726	5.0	1810	13	Q90824	Q90824 gallus gall
42	711.5	4.9	1714	13	Q90995	Q90995 gallus gall
43	675.5	4.6	2019	11	Q64706	Q64706 mus musculu
44	668.5	4.6	1532	13	Q90994	Q90994 gallus gall
45	632.5	4.3	4006	11	Q35452	Q35452 mus musculu

ALIGNMENTS

RESULT 1						
Q9WTS5	Q9WTS5	PRELIMINARY;	PRT;	2764 AA.		
ID	Q9WTS5	PRELIMINARY;	PRT;	2764 AA.		
AC	Q9WTS5;					
DT	01-NOV-1999 (TrEMBLrel. 12, Created)					
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)					
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)					
DE	TEN-M2.					
GN	ODZ2 OR TEN-M2.					
OS	Mus musculus (Mouse).					
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.					
OX	NCBI_TaxID=10090;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-BALB/C; TISSUE-BRAIN;					
RA	Oohashi T., Zhou X., Feng K., Richter B., Moergelin M., Perez M.T.,					
RA	Su W., Chiquet-Ehrismann R., Rauch U., Faessler R.;					
RT	"Mouse Ten-m/Odz is a new family of dimeric type II transmembrane					
RT	proteins expressed in many tissues.";					
RL	J. Cell Biol. 0:0-0(1999).					
DR	EMBL; AB025411; BAA77397.1; -					
DR	HSSP; P35555; IEMN.					
DR	MGD; MGI:1345184; Odz2.					
DR	InterPro; IPR000561; EGF-like.					
DR	InterPro; IPR002049; Laminin_EGF.					
DR	InterPro; IPR001258; NHL.					
DR	Pfam; PF00008; EGF; 4.					
DR	Pfam; PF01436; NHL; 2.					
DR	PRINTS; PR00011; EGF-LAMININ.					
DR	SMART; SM00181; EGF; 5.					
DR	PROSITE; PS00022; EGF_1; UNKNOWN_8.					
DR	PROSITE; PS01186; EGF_2; 7.					
KW	EGF-like domain; Glycoprotein.					
SQ	SEQUENCE 2764 AA; 306464 MW; 73BA3D916D0F0344 CRC64;					

Query Match 98.2%; Score 14316.5; DB 11; Length 2764;  
Best Local Similarity 96.8%; Pred. No. 0;

Matches 2684; Conservative 23; Mismatches 16; Indels 51; Gaps 3;

```
Qy 1 MDVKDRRHSRLTRGRCCKECRYTSSSLDSEDCRVPTQKYSSETLKAAYDHDHSMHYGNR 60
Db 1 MDVKDRRHSRLTRGRCCKECRYTSSSLDSEDCRVPTQKYSSETLKAAYDHDHSMHYGNR 60
Qy 61 VTDLTHRESDEFPQGTNFTLAELGICEPSPHRSYCSDMGILHOGYSLSGSDADSDE 120
Db 61 VTDLTHRESDEFSROGTNFTLAELGICEPSPHRSYCSDMGILHOGYSLSGSDADSDE 120
Qy 121 GMSPEHATRLNGRGIKSRSSGLSRNSALTLTDSNENKSDDENGRPIPTSPSLL 180
Db 121 GMSPEHATRLNGRGIKSRSSGLSRNSALTLTDSNENKSDDENGRPIPTSPSLL 180
Qy 181 PSAQLPSSNPPVSCOMPLLDNSNTSHQIMDTPDEERSPNLYLLRACSGPOQASSGPP 240
Db 181 PSAQLPSSNPPVSCOMPLLDNSNTSHQIMDTPDEERSPNLYLLRACSGPOQASSGPP 240
Qy 241 NHHSOSTLRPLPPPHNHTLSHHSSANSLSNLSLNRSSQIHAPAPAPNDLATTPEVQ 300
Db 241 NHHSOSTLRPLPPPHNHTLSHHSSANSLSNLSLNRSSQIHAPAPAPNDLATTPEVQ 300
Qy 301 LQDSWVLNSNVPLETRHFLFKTSSTPLFSSSSGYPPLTSGVTYPPRLLPRNTEFRK 360
Db 301 LQDSWVLNSNVPLETRHFLFKTSSTPLFSSSSGYPPLTSGVTYPPRLLPRNTEFRK 360
Qy 361 AFKLKPKSKYCSWKCAALSAIAAALLAILLAYFI----- 395
Db 361 AFKLKPKSKYCSWKCAALSAIAAALLAILLAYFIAMHLLGLNWLQAPADGHTFNNGVPT 420
Qy 396 -----VPMLSKNSIDSCEAEVGRVTVQEVPPGVFWRSQIHISQPFUK 439
Db 421 GLPGNDVATVPSSGKVPMSLKNSSIDSCEAEVGRVTVQEVPPGVFWRSQIHISQPFUK 480
Qy 440 FNISLKGKDALFGVYLRRLGPPSHAOYDFMERLDGKEKSWVESPRRRSIQTLVONEAVF 499
Db 481 FNISLKGKDALFGVYLRRLGPPSHAOYDFMERLDGKEKSWVESPRRRSIQTLVONEAVF 540
Qy 500 VOYLDVGLWHLAFYNDGDKEMVSENTVVLDSVQDCPRNCHNGECVSGVCHCFPGFLGA 559
Db 541 VOYLDVGLWHLAFYNDGDKEMVSENTVVLDSVQDCPRNCHNGECVSGVCHCFPGFLGA 600
Qy 560 DCAKAACPVLCSGNGQYSGTKCQCSGWKGACEDVPMNQCIDPCSGHSGSCIDGNCVCSA 619
Db 601 DCAKAACPVLCSGNGQYSGTKCQCSGWKGACEDVPMNQCIDPCSGHSGSCIDGNCVCAA 660
Qy 620 GYKGHECEVDCLDPTCSSHGVVNGECLCPGWGLNCELARVOCPDQC SGHGTYLPT 679
Db 661 GYKGHECEVDCLDPTCSSHGVVNGECLCPGWGLNCELARVOCPDQC SGHGTYLPT 720
Qy 680 GLCSDPNMMDGCDVSVGVSDCGTHGVCIGGACRCEEGWTGAACDQRYCHPRCIEHGTC 739
Db 721 GLCSDPNMMDGCDVSV-VCSVDCGTHGVCIGGACRCEEGWTGAACDQRYCHPRCIEHGTC 779
Qy 740 KDGKCECRGNGEHCITGRQTAGTETDCCPDLCNGNRCRTLQNSWCQVCOTGWRGPC 799
Db 780 KDGKCECRGNGEHCIT-----DCCPDLCNGNRCRTLQNSWCQVCOTGWRGPC 830
Qy 800 NYAMETSCADNKNDEGLVDCLDPDCCLOQACQNSLLCRGSRDPLDIIIOCGOTDWPVK 859
Db 831 NYAMETSCADNKNDEGLVDCLDPDCCLOQACQNSLLCRGSRDPLDIIIOCGOTDWPVK 890
Qy 860 SPYDRKLLAGKDSHTIIPGENPNSLSVLIRGQVVTDTGTPLVGVNVSFYKPKYGT 919
Db 891 SPYDRKLLAGKDSHTIIPGDPNPNSSLSVLIRGQVVTMDGTPLVGVNVSFYKPKYGT 950
Qy 920 ITRQDGTDLIANGASULTLHFERAPPMSQERTVMLPWNFSFYAMDTLVNKTTEENIPSCD 979
Db 951 ITRQDGTDLIANGASULTLHFERAPPMSQERTVMLPWNFSFYAMDTLVNKTTEENIPSCD 1010
Qy 980 LSGFVRPDPPIIISSPLSTFESAACQNPVTPQVTLHBEIEPLPGSNVKLRYLSSRTAGYK 1039
Db 1011 LSGFVRPDPPIIISSPLSTFESAASPNLPVTPQVTLHBEIEPLPGSNVKLRYLSSRTAGYK 1070
```

```
Qy 1040 SLLKITMTOSTVPLNLIRVHLMVAVEGHLFOKSFQASPNLASTFTIWDKTDAYQORVYGLS 1099
Db 1071 SLLKITMTOSTVPLNLIRVHLMVAVEGHLFOKSFQASPNLAYTFIWDKTDAYQORVYGLS 1130
Qy 1100 DAVSVGFYEYETCPSLILWEKRTALLQGFELDPNSLGGMSLDKHHILNVKSGILHKGTE 1159
Db 1131 DAVSVGFYEYETCPSLILWEKRTALLQGFELDPNSLGGMSLDKHHILNVKSGILHKGTE 1190
Qy 1160 NOFLTQOPALITTSIMGNRRRSISPCSCNGLAEGNKLAPALAVAGIDGSLVGVDFNYIR 1219
Db 1191 NOFLTQOPALITTSIMGNRRRSISPCSCNGLAEGNKLAPALAVAGIDGSLVGVDFNYIR 1250
Qy 1220 RIFPSRNTVTSILELRNKEFKHNNPAHKYYLAVDPVSGSLYVSDTNSRRIYVKSLSGTFK 1279
Db 1251 RIFPSRNTVTSILELRNKEFKHNSPGHKYYLAVDPVSGSLYVSDTNSRRIYVKSLSGAK 1310
Qy 1280 DIAGNSEVVAGTGEQCLPFDEARCGDGGKAIDATLMSPRGIAVKNGLMYFVDATMIRKV 1339
Db 1311 DIAGNSEVVAGTGEQCLPFDEARCGDGGKAVDATLMSPRGIAVKNGLMYFVDATMIRKV 1370
Qy 1340 DONGIISTLLGSNDLTAVRPLSCDSSMOVAOVRLEWPTDLAVNPMONSILVLENNVILRI 1399
Db 1371 DONGIISTLLGSNDLTAVRPLSCDSSMOVAOVRLEWPTDLAVNPMONSILVLENNVILRI 1430
Qy 1400 TENHOVSIITAGRPMHCQVPGIDYSLSKLAHSALESASAIASHTGVLVITETDEKKINR 1459
Db 1431 TENHOVSIITAGRPMHCQVPGIDYSLSKLAHSALESASAIASHTGVLVITETDEKKINR 1490
Qy 1460 LRQVTTNGEICLLAGAAASDCCKNDVNCNCSGDDAYATDAILNSPSSSLAVAPDGTIYTA 1519
Db 1491 LRQVTTNGEICLLAGAAASDCCKNDVNCNCSGDDAYATDAILNSPSSSLAVAPDGTIYTA 1550
Qy 1520 DLGNTRIRAVSKNKPVLNAFNOYEAAASPGOEQLYVFNADGIIHOYTSLVTVGYLVNFTYS 1579
Db 1551 DLGNTRIRAVSKNKPVLNAFNOYEAAASPGOEQLYVFNADGIIHOYTSLVTVGYLVNFTYS 1610
Qy 1580 TDNDVTELDINNGNSLKIIRRDSSGMPRHLLMPDNOIITLTVTGNTGGLKVYVSTONLEGLM 1639
Db 1611 ADNDVTELDINNGNSLKIIRRDSSGMPRHLLMPDNOIITLTVTGNTGGLKAVSTONLEGLM 1670
Qy 1640 TYDGNLTGLLTKSDGTGTTTFYDDBEGRLTNVTRPTGVVTVSLHREMEKSIITIDIENSR 1699
Db 1671 TYDGNLTGLLTKSDGTGTTTFYDDBEGRLTNVTRPTGVVTVSLHREMEKSIITIDIENSR 1730
Qy 1700 DDDVTVITNLSSEASVYTVVQDVQRNSYQLCNNGTLRVMYANGMGSFSEPHVLGTTT 1759
Db 1731 DDDVTVITNLSSEASVYTVVQDVQRNSYQLCNNGTLRVMYANGMAVSPHSEPHVLGTTT 1790
Qy 1760 PTIGRCNLSLPMENGLNLSIEWRLKEQIKGKVTIFGRKLVRHGRNLLSIDYDORNIIRTEKI 1819
Db 1791 PTIGRCNLSLPMENGLNLSIEWRLKEQIKGKVTIFGRKLVRHGRNLLSIDYDORNIIRTEKI 1850
Qy 1820 YDDBHRTLRILYIDOVGRPFLWLPSSGLAAVNVSYFENGRLAGLORGAMSERTDIDKGR 1879
Db 1851 YDDBHRTLRILYIDOVGRPFLWLPSSGLAAVNVSYFENGRLAGLORGAMSERTDIDKGR 1910
Qy 1880 IVSRMFADCKVMSYSYLDKSMVLLLSQSRQYIFEDYDSSDRLLAVTMPSVARHSMSTHTSI 1939
Db 1911 IVSRMFADCKVMSYSYLDKSMVLLLSQSRQYIFEDYDSSDRLLAVTMPSVARHSMSTHTSI 1970
Qy 1940 GYIRNIYNPPESNASVIEDYSDDGRIKLTSTFLGTGRQVFKYKYGKLSKJSEIYVYSTAVTF 1999
Db 1971 GYIRNIYNPPESNASVIEDYSDDGRIKLTSTFLGTGRQVFKYKYGKLSKJSEIYVYSTAVTF 2030
Qy 2000 GYDETTGVKLKMNVLQSGGFSCTIRYRKTCPLVDKQIYRFSEBEGMKNARFDYTHONSERI 2059
Db 2031 GYDETTGVKLKMNVLQSGGFSCTIRYRKTCPLVDKQIYRFSEBEGMKNARFDYTHONSERI 2090
Qy 2060 ASIKPVISETPLVDLYRYDEISGKVEHFGKFGVYIYDINOIITAVMTLSKHFTDHGRI 2119
Db 2091 ASIKPVISETPLVDLYRYDEISGKVEHFGKFGVYIYDINOIITAVMTLSKHFTDHGRI 2150
```

Qy 2120 KEVQYEMFRSLMYMTVOYDSMGVRVIRKELKGLGYANTTKYTYDYGQQLQSVAVVNDP 2179  
Db 2151 KEVQYEMFRSLMYMTVOYDSMGVRVIRKELKGLGYANTTKYTYDYGQQLQSVAVVNDP 2210  
Qy 2180 TWRSYDLNGLHLLNPGNSVRLMPLRYDLRDRITRGLDVOYKIDDDGYLCORGSDIFEY 2239  
Db 2211 TWRSYDLNGLHLLNPGNSARLMLPLRYDLRDRITRGLDVOYKIDDDGYLCORGSDIFEY 2270  
Qy 2240 NSKGLLFRAYNKASGWSQVRYDCVGRASYKTNLGHHLQYFYSDLNHPNTRITHVYNSN 2299  
Db 2271 NSKGLLFRAYNKASGWSQVRYDCVGRASYKTNLGHHLQYFYSDLNHPNTRITHVYNSN 2330  
Qy 2300 SEITSLYDLDQHLFAMESSGEYVYASDNTGTPPLAVFSINGLMIKOLQYATYGEIYD 2359  
Db 2331 SEITSLYDLDQHLFAMESSGEYVYASDNTGTPPLAVFSINGLMIKOLQYATYGEIYD 2390  
Qy 2360 SNPDFQMVIGPHGLYDPLTKLVHFTQDYDVLGRWTSPTYTMWKNVGRKPEPFLNLYM 2419  
Db 2391 SNPDFQMVIGPHGLYDPLTKLVHFTQDYDVLGRWTSPTYTMWKNVGRKPEPFLNLYM 2450  
Qy 2420 KSNPLSSELDLKNYVDVKSWMVYFQLSNIIPGPRAKMYFVPPPYELSESQASENG 2479  
Db 2451 KSNPLSSELDLKNYVDVKSWMVYFQLSNIIPGPRAKMYFVPPPYELSESQASENG 2510  
Qy 2480 QLITGVQOOTERRHNOAFMALEGQVITKHLASIREKAGHWFATTPPIIGKIMFAIKR 2539  
Db 2511 QLITGVQOOTERRHNOAFMALEGQVITKHLASIREKAGHWFATTPPIIGKIMFAIKR 2570  
Qy 2540 VTTGVSSITASEDSRKVASLVNNAYLDKMHYSIEGKTHYFVKIGSADGDLVLTGTTGR 2599  
Db 2571 VTTGVSSITASEDSRKVASLVNNAYLDKMHYSIEGKTHYFVKIGSADGDLVLTGTTGR 2630  
Qy 2600 KVLSEGVNVTVSQPTLLVNGTRFTNIEFOYSTLLSIRYGLTPDPTLDEKARVLQDAR 2659  
Db 2631 KVLSEGVNVTVSQPTLLVNGTRFTNIEFOYSTLLSIRYGLTPDPTLDEKARVLQDAR 2690  
Qy 2660 QRALGTAWAKEQQKARDGREGSLWTEGEKQOLLSTGRVQYEGYVYLPVEQYPELADSS 2719  
Db 2691 QRALGTAWAKEQQKARDGREGSLWTEGEKQOLLSTGRVQYEGYVYLPVEQYPELADSS 2750  
Qy 2720 SNIQFLRQNMGR 2733  
Db 2751 SNIQFLRQNMGR 2764

RESULT 2  
Q9R1K2 PRELIMINARY; PRT; 2765 AA.  
Q9R1K2; 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE NEURESTIN ALPHA.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
SEQUENCE FROM N.A.  
PC STRAIN-SPRAGUE DAWLEY; TISSUE-OLFACTORY BULB;  
RX MEDLINE=9330226; PubMed=10419693;  
RA Ozaki J.M., Firestein S.;  
RT "Neurestin: putative transmembrane molecule implicated in neuronal development.";  
RL Dev. Biol. 212:165-181(1999).  
DR EMBL; AF086607; AAD47393.1; -;  
DR HSSP; P00750; ITPG.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR002049; Laminin\_EGF.  
DR InterPro; IPR001258; NHL.  
DR Pfam; PF000008; EGF; 4.  
DR Pfam; PF01436; NHL; 2.  
DR PRINTS; PR00011; EGF/LAMININ.

DR SMART; SMO0181; EGF; 5.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_8.  
DR PROSITE; PS01186; EGF\_2; 7.  
KW EGF-like domain; Glycoprotein.  
SQ SEQUENCE 2765 AA; 306568 MW; 6748D70D5FFD8F0E CRC64;  
  
Query Match 98.0%; Score 14291; DB 11; Length 2765;  
Best Local Similarity 96.5%; Pred. No. 0;  
Matches 2676; Conservative 30; Mismatches 18; Indels 50; Gaps 2;  
  
Qy 1 MDVKDRHRSLTRGRCCKEYRYSSTLSDSDCRVPTOKSYSSSETLKAYDHDSDMHYGNR 60  
Db 1 MDVKDRHRSLTRGRCCKEYRYSSTLSDSDCRVPTOKSYSSSETLKAYDHDSDMHYGNR 60  
Qy 61 VTDLIHRESDEFPRQGFNFTLAEIGICEPSPHRSGYCDMGILHGGYSLTSGSDADSDTE 120  
Db 61 VTDLIHRESDEFPRQGFNFTLAEIGICEPSPHRSGYCDMGILHGGYSLTSGSDADSDTE 120  
Qy 121 GMSPEHAIRLWGRGKSRSSGLSSRENSALTLTDSNENKSDDENGRPIPTTSSPLL 180  
Db 121 GMSPEHAIRLWGRGKSRSSGLSSRENSALTLTDSNENKSDDENGRPIPTTSSPLL 180  
Qy 181 PSAQLPSSHNPVVSQMPPLDSDNTSHQIMDTPDEEFSPNSYLLRACSGPQQASSGPP 240  
Db 181 PSAQLPSSHNPVVSQMPPLDSDNTSHQIMDTPDEEFSPNSYLLRACSGPQQASSGPP 240  
Qy 241 NHHQSOTLRPLPPPHNHTLSHHSSANSLSNRSLTNRROIHAPAPAPNDLATTPEVQ 300  
Db 241 NHHQSOTLRPLPPPHNHTLSHHSSANSLSNRSLTNRROIHAPAPAPNDLATTPEVQ 300  
Qy 301 LQDSWLVNSVPLETRHFLFKTSSTGSPFLFSSSSPGYPLTSGTYVTPPRLLPNTFSRK 360  
Db 301 LQDSWLVNSVPLETRHFLFKTSSTGSPFLFSSSSPGYPLTSGTYVTPPRLLPNTFSRK 360  
Qy 361 AFKLKSKYCSWKCAALSATAALLAILLAYFI----- 395  
Db 361 AFKLKSKYCSWKCAALSATAALLAILLAYFIAMHLLGLNWLQPADGHTFNNGVRT 420  
Qy 396 -----VPWSLKNSIDSGEAEVGRRTVQVPPGVFWRSQTHISQPF 439  
Db 421 GLPNDVDVATVPSSGKVPWSLKNSIDSGEAEVGRRTVQVPPGVFWRSQTHISQPF 480  
Qy 440 FNISLGDALFGVYIRRLPSPHAQYDFMERLDGKEMSVESPRERSIQTLYQNEAVF 499  
Db 481 FNISLGDALFGVYIRRLPSPHAQYDFMERLDGKEMSVESPRERSIQTLYQNEAVF 540  
Qy 500 VOYLDVGLWHLAFYNDGKDEKEMSVFNTVLDSDVQDCPRNCHNGECVSGVCHCFPGFLA 559  
Db 541 VOYLDVGLWHLAFYNDGKDEKEMSVFNTVLDSDVQDCPRNCHNGECVSGVCHCFPGFLA 600  
Qy 560 DCAKACPVLCSGNGQYSGTCQCYSGWKGAECVPMNQCIDPSCGGHSCIDGNCVCSA 619  
Db 601 DCAKACPVLCSGNGQYSGTCQCYSGWKGAECVPMNQCIDPSCGGHSCIDGNCVCSA 660  
Qy 620 GYKGEHCEVDCLDPTCSSHGCVNGECLSPGNGLNCCLARVQCPCQSGHGTYLPDT 679  
Db 661 GYKGEHCEVDCLDPTCSSHGCVNGECLSPGNGLNCCLARVQCPCQSGHGTYLPDT 720  
Qy 680 GLCSCDPNMGWPCDSEVSCVSDCGTHGVCIGGACRCEGWTGAACDQVCHPRCIEHCTC 739  
Db 721 GLCNCDPNMGWPCDSEVSCVSDCGTHGVCIGGACRCEGWTGAACDQVCHPRCIEHCTC 780  
Qy 740 KDGKCECEGNGHGHCTTGRTAGTETDGCPLDNGNCRCTLGNSWQCVQCTQWRGPGC 799  
Db 781 KDGKCECEGNGHGHCTI-----DGDPLDNGNCRCTLGNSWQCVQCTQWRGPGC 831  
Qy 800 NVAMETSCADKNDEGDLVCLDPCCLQSAQNSLLCRGRDPLDIIOGQTDWPAVK 859  
Db 832 NVAMETSCADKNDEGDLVCLDPCCLQSAQNSLLCRGRDPLDIIOGQTDWPAVK 891  
Qy 860 SFYDRILKAGKSTHIIPGENPENSLSVLRQVVTDTGTLPLGVNVSFVKPKYGT 919  
Db 919 SFYDRILKAGKSTHIIPGENPENSLSVLRQVVTDTGTLPLGVNVSFVKPKYGT 919

Db 892 SFYDRIKLAGKSTHIIIPGDNPFNSLSLIRGQVVTGDTPLVGVNVSFKYKPYGVT 951  
Qy 920 ITRQDGTFLIANGASLTFLHFERAPFMQERTVWLPWNSFYAMDTLVNMTKEENSIPSCD 979  
Db 952 ITRQDGTFLIANGASLTFLHFERAPFMQERTVWLPWNSFYAMDTLVNMTKEENSIPSCD 1011  
Qy 980 LSGFVRPDPPIIISSPLSTFFSAAPQNPVPTQVLHHEIELPGSNVKLYLSSRTAGYK 1039  
Db 1012 LSGFVRPDPPIIISSPLSTFFSAAPNPVPTQVLHHEIELPGTNVKLYLSSRTAGYK 1071  
Qy 1040 SILKTTMTQSTVPLNLRVHLVAVVEGHLFQKSFQASPNLSTFFWDKTDAYGQRYGLS 1099  
Db 1072 SILKTTMTQSTVPLNLRVHLVAVVEGHLFQKSFQASPNLAYTFFIWDKTDAYGQRYGLS 1131  
Qy 1100 DAVSVGFYEYETCPSLILWEKRTALLOQFELPDSNLGGHSLDKHHTLVNKSILKHGTGE 1159  
Db 1132 DAVSVGFYEYETCPSLILWEKRTALLOQFELPDSNLGGHSLDKHHTLVNKSILKHGTGE 1191  
Qy 1160 NOFLTQOPAIITISINGNRRRRISPCSCNGLAEGNKLAPVALAVIGDGLSYGVDFNYIR 1219  
Db 1192 NOFLTQOPAIITISINGNRRRRISPCSCNGLAEGNKLAPVALAVIGDGLSYGVDFNYIR 1251  
Qy 1220 RIFPGRNVTISILELRNKEPKHNNPAHKYVYLAQVPSGLSYVSDTNSRRIRYRVKSLSGTK 1279  
Db 1252 RIFPGRNVTISILELRNKEPKHNSPGHKYVYLAQVPTGSLYVSDTNSRRIRYRVKSLSGAK 1311  
Qy 1280 DLAGNSEVVAGTGEQCLPFDEARCGDGGKRAIDATLMSPRGIAVDKNGLMYFVDATMIRKV 1339  
Db 1312 DLAGNSEVVAGTGEQCLPFDEARCGDGGKAVDATLMSPRGIAVDKNGLMYFVDATMIRKV 1371  
Qy 1340 DQNGIISTLLGNDLTAVRPLSCDSSMDVAQVRLWPTDLAVNPMDNSLYVLENNVILRI 1399  
Db 1372 DQNGIISTLLGNDLTAVRPLSCDSSMDVAQVRLWPTDLAVNPMDNSLYVLENNVILRI 1431  
Qy 1400 TENHOVSIITAGPMHCQVPGIDYLSKLAHSALESASAIATSHGTGLYITETDEKKINR 1459  
Db 1432 TENHOVSIITAGPMHCQVPGIDYLSKLAHSALESASAIATSHGTGLYITETDEKKINR 1491  
Qy 1460 LRQVTTNGEICLLAGAAASDCCKNDVNCNCSGDDAYATDAILNSPSSLAVALPDGTIYIA 1519  
Db 1492 LRQVTTNGEICLLAGAAASDCCKNDVNCNCSGDDAYATDAILNSPSSLAVALPDGTIYIA 1551  
Qy 1520 DLGNTRIRAVSNKPKVLNAFNQYEAASPGQEELYFVNADGIIHOYTIVSLVTGEYLNFTYS 1579  
Db 1552 DLGNTRIRAVSNKPKVLNAFNQYEAASPGQEELYFVNADGIIHOYTIVSLVTGEYLNFTYS 1611  
Qy 1580 TDNDVTELDNNGNSLKIRRDSSGMPRHLLMPDNQIITLVCTNGGLKVWSTONLELGLM 1639  
Db 1612 ADNDVTELDNNGNSLKIRRDSSGMPRHLLMPDNQIITLVCTNGGLKAVSTONLELGLM 1671  
Qy 1640 TYDGNLTGLLATKSDGTWTFYDYDHEGRLTNVTRPTGVVTSLHREMEKSIITIDIENSR 1699  
Db 1672 TYDGNLTGLLATKSDGTWTFYDYDHEGRLTNVTRPTGVVTSLHREMEKSIITIDIENSR 1731  
Qy 1700 DDDVTIVTNLSSVEASYTVVQOVNRNSYQLCNGNGLRVNANGMGSIFHSEPHVLAGTIT 1759  
Db 1732 DNDVTIVTNLSSVEASYTVVQOVNRNSYQLCNGNGLRVNANGMGSIFHSEPHVLAGTIT 1791  
Qy 1760 PTIGRCNTSLPENGNSLIEWRLKEQIKGVTIFGRKLVRHGRNLLSIDYDRNIRTEKI 1819  
Db 1792 PTIGRCNTSLPENGNSLIEWRLKEQIKGVTIFGRKLVRHGRNLLSIDYDRNIRTEKI 1851  
Qy 1820 YDHRKFTLRIIYDQVGRPFLWLPSSGLAAVNVSYFFNGRLAGLQRGAMSERTIDKQGR 1879  
Db 1852 YDHRKFTLRIIYDQVGRPFLWLPSSGLAAVNVSYFFNGRLAGLQRGAMSERTIDKQGR 1911  
Qy 1880 IYSRWFADGKVNYSYLDKSWLLVLSQSQRYIFEYDSSDRLLAVTMPVSARHSMSTHTSI 1939  
Db 1912 IYSRWFADGKVNYSYLDKSWLLVLSQSQRYIFEYDSSDRLLAVTMPVSARHSMSTHTSI 1971  
Qy 1940 GYIRINYNPPESNASVIFDYSDGRILKTSLFGTGRQVFKYKGLSKLSLSEIYVDSVTATVF 1999  
Db 1972 GYIRINYNPPESNASVIFDYSDGRILKTSLFGTGRQVFKYKGLSKLSLSEIYVDSVTATVF 2031

Qy 2000 GYDETTGVLKMNVLQSGGFSCTIRYRKIGPLVDKQIYREFSEEGMYNARFDYTHDNSFRI 2059  
Db 2032 GYDETTGVLKMNVLQSGGFSCTIRYRKIGPLVDKQIYREFSEEGMYNARFDYTHDNSFRI 2091  
Qy 2060 ASIKPVISETPLVDLYRYDEISGKVEHFGKFGVYIYDINQIITAVMTLSKHFDTHGRI 2119  
Db 2092 ASIKPVISETPLVDLYRYDEISGKVEHFGKFGVYIYDINQIITAVMTLSKHFDTHGRI 2151  
Qy 2120 KEVOYEMERSLAMYMTVQYDSMGRVIRKELKGPVANTTKYTDYDGGLOLOSVAANDRP 2179  
Db 2152 KEVOYEMERSLAMYMTVQYDSMGRVIRKELKGPVANTTKYTDYDGGLOLOSVAANDRP 2211  
Qy 2180 TWRYSYDLNGLNHLNPGNSVRLMPLRYDLRDRITRLGSDVQYKIDDDGVLCOGSDIFEY 2239  
Db 2212 TWRYSYDLNGLNHLNPGNSVRLMPLRYDLRDRITRLGSDVQYKIDDDGVLCOGSDIFEY 2271  
Qy 2240 NSKGLLTRYANKASGWSVOYRYDGVGRASRYKTNLGHLOFYFYSDLHNPTRTHYHNISN 2299  
Db 2272 NSKGLLTRYANKASGWSVOYRYDGVGRASRYKTNLGHLOFYFYSDLHNPTRTHYHNISN 2331  
Qy 2300 SEITSLYDLQGLHFLAMESSEGEYVYVADNTGTPLAVFSINGLMIKOLQYTAIGEYVD 2359  
Db 2332 SEITSLYDLQGLHFLAMESSEGEYVYVADNTGTPLAVFSINGLMIKOLQYTAIGEYVD 2391  
Qy 2360 SNPDFQMVIGFHGGLYDPLTKLVHFTQDYDVLGRWTSPODYTMKNVNGKEPAPFNLYMF 2419  
Db 2392 SNPDFQMVIGFHGGLYDPLTKLVHFTQDYDVLGRWTSPODYTMKNVNGKEPAPFNLYMF 2451  
Qy 2420 KSNPLSELDDLKNVYTDVKSMLVMFGQLSNIIIPGFPRAKMYFVPPPYELSESASENG 2479  
Db 2452 KSNPLSELDDLKNVYTDVKSMLVMFGQLSNIIIPGFPRAKMYFVPPPYELSESASENG 2511  
Qy 2480 QLITGVQOOTERRHNOAFMALEGOVITTKLHASIREKAGHWFATTPIIGKGMFAIKGR 2539  
Db 2512 QLITGVQOOTERRHNOAFMALEGOVITTKLHASIREKAGHWFATTPIIGKGMFAIKGR 2571  
Qy 2540 VTTGVSSIASSEDSRVSALNNAYYLDKMHYSIEGKDTHYFVKIGSAGDLVTLGTTIGR 2599  
Db 2572 VTTGVSSIASSEDSRVSALNNAYYLDKMHYSIEGKDTHYFVKIGSAGDLVTLGTTIGR 2631  
Qy 2600 KYLESSVNVTVSQPTLLVNGRTRRTNIEFOYSTLLSTRYGLTPTDLDDEKARVLDQAR 2659  
Db 2632 KYLESSVNVTVSQPTLLVNGRTRRTNIEFOYSTLLSTRYGLTPTDLDDEKARVLDQAR 2691  
Qy 2660 ORALCTAWAKEQOKARDREGSRLWTEGEKQOLLSTGRVQVGYEYVLPVEQYPELADSS 2719  
Db 2692 ORALCTAWAKEQOKARDREGSRLWTEGEKQOLLSTGRVQVGYEYVLPVEQYPELADSS 2751  
Qy 2720 SNIQFLRONEMGRK 2733  
Db 2752 SNIQFLRONEMGRK 2765

## RESULT 3

QSDERS  
ID Q9DER5 PRELIMINARY; PRT: 2802 AA.  
AC Q9DER5;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE TENEURIN-2 (FRAGMENT).  
GN TEN2.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20581705; PubMed=11146505;  
RA Tucker R.P., Chiquet-Ehrismann R., Chevrion M., Martin D., Hall R.J.,  
RA Rubin B.P.;

"Teneurin-2 is expressed in tissues that regulate limb and somite  
RT pattern formation and is induced in vitro and in situ by FGF8.";  
RL Dev. Dyn. 220:27-39(2001).  
DR EMBL; AJ279031; CAC09416.1; -.  
DR HSPB; P00750.1TPG.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR002049; Laminin\_EGF.  
DR Pfam; PF00008; EGF; 5.  
DR PRINTS; PR00011; EGF\_LAMININ.  
DR SMART; SM00181; EGF; 7.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
FT NON\_TER 2802 2802  
SQ SEQUENCE 2802 AA; 310745 MW; B1FBC2C84EDFA4B3 CRC64;

Query Match	95.68;	Score 13934.5;	DB 13;	Length 2802;
Best Local Similarity	92.38;	Pred. No. 0;		
atches 2586;	Conservative 95;	Mismatches 52;	Indels 69;	Gaps 2;
1 MDVKRRHRSLTRGRCGKECRYTSSSLDSEDCRVPTOKSYSSSETELKAYDHSRHHYGNR 60				
1 MDIKDRHRSLTRGRCGKECRYTSSSLDSEDCRVPAQKSYSSSETELKAYGHDTRHHYGNR 60				
61 VTDLTHRSDEFFPROGNTNFTLAELGICEPSPHRSGYCSMDGILHOGYSLSTGSDADSDE 120				
61 VSDLVHRSDEFFPROGNTNFTLAELGICEPSPHRSGYCSMDGILHOGYSLSTGSDADSDE 120				
121 GGMSPHAIRLWGRIKSRSSGLSSRENSALTLDSDNENKSDDENG----- 168				
121 GGMSPHAIRLWGRIKSRSSGLSSRENSALTLDSDNENKSDDENGFHTLSEKLDKR 180				
169 -----RIPPTSSPLPSAQLPSSHNPVPVSCOMPLLDNSTHQIMDT 212				
181 QTSWQOLAETKNSLRIRPTSSSLPSAQLPSSHNPVPVSCOMPLLDNSTHQIMDT 240				
213 NPDEFSNYSLLRACSGPQAASSGPPNHHHSQSTLRPLPPPHNHTLSHHSSANSNR 272				
241 NPDEFSNYSLLRACSGPQAASSGPPNHHHSQSTLRPLPPPHNHTLSHHSSANSNR 300				
273 NSLTNRROIHAPAPANDLATTPEVQLODSWLNPNVPLETRHFLFKTSSTGTPLESS 332				
301 NSLTNRROIHAPAPANDLATTPEVQLODSWLNPNVPLETRHFLFKTSSTGTPLESS 360				
333 SSPGYPLTSGTYTTPPRLPNTESRKAFLKPKSKYCSWKCAALSAIAAALLAILLA 392				
361 SSPGYPLTSGTYTTPPRLPNTESRKAFLKPKSKYCSWKCAALSAIAAALLAILLA 420				
393 YPIV-----PWSLKNSSIDSGAE 411				
421 YPIAMHLLGLNWLQPADGHTFNSGLRPGAAGAEAGAAAPAGRGPVWTRNSSIDSGETE 480				
412 VGRRYTQEVPPGVFWRISOIHISOPQFLKFNISLGDALFGVYIRGLPSPAQYDFMERL 471				
481 VGRKYTQEVPPGVFWRISOIHISOPQFLKFNISLGDALFGVYIRGLPSPAQYDFMERL 540				
472 DGKEKWSVESPRRRSIQTLVQNEAVFQYLDVGLWHLAFYNDGKDEKEMVSFNTVWLDS 531				
541 DGKEKWSVESPRRRSIQTLVQNEAVFQYLDVGLWHLAFYNDGKDEKEMVSFNTVWLDS 600				
532 VODCPRNCHNGECVSGVCHCFPGFLGADCAKAAAPVLCGNGQYVKTCOCYSGWKGA 591				
601 VODCPRNCHNGECVSGVCHCFPGFLGADCAKAAAPVLCGNGQYVKTCOCYSGWKGA 660				
592 CDVPMNOCIDPSCGGHSGCIDGNCVCSAGYKGEHCEEVDCDPTCSSHGVCVNGECLCSP 651				
661 CDVPIISQCIDPSCGGHSGCIDGNCVCSAGYKGEHCEEVDCDPTCSSHGVCVNGECLCSP 720				
652 GWGLNCELARVQCDDQSGHGTLYLPDTGLCSDCPNMMGPDGSEVVCVSDCTHGCVCIGG 711				
721 GWGINCELARVQCDDQSGHGTLYLPDTGLCSDCPNMMGPDGSEVVCVSDCTHGCVCIGG 780				
712 ACRCEEGWTGAACDQVRVCHPRCIEHGTCKDKGKCECREGNGEHCTIGROTAGTETDGC 771				

Db	781	ACRCEEGWTGAACDQVRVCHPRCIEHGTCKDKGKCECREGNGEHCTIGROTAGTETDGC 840
Qy	772	LCNGNGRCTLGQNSWQVCQGTGWRGPGCNVAMETSCADKNKNEGDLVDCLDPPCCQLQSA 831
Db	841	LCNGNGRCTLGQNSWQVCQGTGWRGPGCNVAMETSCADKNKNEGDLVDCLDPPCCQLQST 900
Qy	832	QNSLLCRGSRDPLDIIOQGGTDMPAKVSFVDRIKLLAGKDSHTIIPGENPNSSLVSLI 891
Db	901	QNSLLCRGSRDPLDIIOQSHSGSPAVASFVDRIKLLVKGKDSHTIIPGENPNSSLVSLI 960
Qy	892	RQGVVTTDGTPLVGVNVSVFYKPKYGTITRODGTFDLIANGASLTLHFERRAPMSQER 951
Db	961	RQGVVTTDGTPLVGVNVSVFYKPKYGTITRODGMFDFLVANGSSSLTLHFERRAPMSQER 1020
Qy	952	TWLPWNFSFYAMDTLVMKTEENSIPSCDLSGFVRPDPPIIISPLSTFFSAAPGNPIYPE 1011
Db	1021	TWLPWNFSFYAMDTLVMKTEENSIPSCDLSGFVRPDPPIIISPLSTFFSDAPGRNPIYPE 1080
Qy	1012	TQVLHEETELPGSNVKKRLYSRTAGYKSLKLTMTQSTVPLNLRVHLMVAVEGHLFQK 1071
Db	1081	TQVLHEETELPGSSIKLIYLSRTAGYKSLKIINTQSLVPLNLIKVLHLMVAVEGHLFQK 1140
Qy	1072	SFQASPNLASTFIWDKTDAYGORVYGLSDAVVSVGFYETCPSLTLWEKRTALLOGFELD 1131
Db	1141	SFLASPNLAYTFIWDKTDAYQKVYGLSDAVVSVGFYETCPSLTLWEKRTALLOGFELD 1200
Qy	1132	PSNLGGWSLDDKHHILNVKSGILHKGTGENOFLOTPAIIITSMGNRRRSISCPSCNGLA 1191
Db	1201	PSNLGGWSLDDKHHILNVKSGILHKGTGENOFLOTPAIIITSMGNRRRSISCPSCNGLA 1260
Qy	1192	EGNKLLAPVALAVGIDGSLYVGDFNYIRRIIPSRNVTSILELRNKEFKHNNPAHKYYLA 1251
Db	1261	EGNKLLAPVALAVGIDGSLYVGDFNYIRRIIPSRNVTSILELRNKEFKHNNPAHKYYLA 1320
Qy	1252	VDPVSGSLYVSDTNSRRIRYRVKSLGTDKDLAGNSVWAGTGEQCLPFEARCGDGGKAID 1311
Db	1321	VDPVSGSLYVSDTNSRRIRYRVKSLGTDKDLAGNSVWAGTGEQCLPFEARCGDGGKAID 1380
Qy	1312	ATLMSPRGIADVKNGLMYFVDATMIRKVDQNGIISTLGSNDLTAVRPLSCDSSMDVAQV 1371
Db	1381	ATLMSPRGIADVKNGLMYFVDATMIRKVDQNGIISTLGSNDLTAVRPLSCDSSMDVAQV 1440
Qy	1372	RLEWPTDLAVPMONSLYVLNNVILRITENHQVSIAGRPMHCQVPGIDYSLSKLAHS 1431
Db	1441	RLEWPTDLAVPMONSLYVLNNVILRITENHQVSIAGRPMHCQVPGIDYSLSKLAHS 1500
Qy	1432	ALESASATAISHTGVLYITETDEKKINRLQVTTNGBEICLLAGAAASDCCKNDVNCNYS 1491
Db	1501	ALESASATAISHTGVLYITETDEKKINRLQVTTNGBEICLLAGAAASDCCKNDVNCNYS 1560
Qy	1492	GDDAYATDAILNSPSSLAAPDGTIYIADLGNIRIRAVSKNKPVLNAPNOYEAAASPGQE 1551
Db	1561	GDDAYATDAILNSPSSLAAPDGTIYIADLGNIRIRAVSKNKPVLNAPNOYEAAASPGQE 1620
Qy	1552	LYVFNADGIHOYTVSLVTGEYLYNFTYTDNDVTEDIDNNGNSLIRKRDSSGMPRHLLMP 1611
Db	1621	LYVFNADGIHOYTVSLVTGEYLYNFTYTDNDVTEDIDNNGNSLIRKRDSSGMPRHLLMP 1680
Qy	1612	DNQIITLVGTVNGGLKLVYSTQNLGLMTYDGTGTLATKSDGTGTTFTFYDHDHGRLTN 1671
Db	1681	DNQIITLVGTVNGGLKLVYSTQNLGLMTYDGTGTLATKSDGTGTTFTFYDHDHGRLTN 1740
Qy	1672	VTRPTGVVTSLHREMEKSIITTDIENSRDDVTVTNLTNLSVVEASTVVDQDVNRNSYQLCN 1731
Db	1741	VTRPTGVVTSLHREMEKSIITTDIENSRDDVTVTNLTNLSVVEASTVVDQDVNRNSYQLCN 1800
Qy	1732	NGTLRMVYANGMISFHPSEPHVLAQITPTIGRCNISLPMENGLNSIEWRLRKEQIKGV 1791
Db	1801	NGTLRMVYANGMISFHPSEPHVLAQITPTIGRCNISLPMENGLNSIEWRLRKEQIKGV 1860
Qy	1792	TIFGRKRLVHGRNLLSIDYDRNIRTEKIYDHRKFTLRIIYDQVGRPFLWLPSSGLA 1851

Db	1861	TVFGRKLVRHGRNLLSIDYDRNRTRTEKIIDDDHRKFTLRIIYDQIGRPFLLWLPSSGLAAVN	1920
QY	1852	VSFFNGRLAGLQRGAMSERTDIDKQGRIVSRMFADGKVMYSYLDKSMVLLLSQSORQYI	1911
Db	1921	VSFFNGRLAGLQRGAMSERTDIDKQGRIVSRMFADGKVMYSYIYLEKSMVLLLSQSORQYI	1980
QY	1912	FEYDSSDRLLAVTMPSPVARHSMSTHTSIGYIRNIYNPPESNASVIFDYSDDGRILKTSFL	1971
Db	1981	FEYDSSDRLLAVTMPSPVARHSMSTHTSVGIRNIYNPPESNASVIFDYSDDGRILKTSFL	2040
QY	1972	GTCRQVFFYKGLKSLSEIYVDSTAVTFYDETTGVLKWNVLQSGGFSCTIRYRKIGPLV	2031
Db	2041	GTCRQVFFYKGLKSLSEIYVDSTAVTFYDETTGVLKWNVLQSGGFSCTIRYRKIGPLV	2100
QY	2032	DKQIYRFSEGMVNARFDVTHDNSPRIASIKPVISETPLPVDLYRYDEISGKVEHFGKF	2091
Db	2101	DKQIYRFSEGMVNARFDVTHDNSPRIASIKPVISETPLPVDLYRYDEISGKVEHFGKF	2160
QY	2092	GVIIYDQINQIITAVMTLSKHFDTHGRIRKEVQYEMFRSLMYMTVOYDSMGRVYKRELKL	2151
Db	2161	GVIIYDQINQIITAVMTLSKHFDTHGRIRKEVQYEMFRSLMYMTVOYDSMGRVYKRELKL	2220
QY	2152	GPYANTTKTYDDGQGLQSVAVNDPRTWRYSYDNLGNLHLLNPGNSVRLMPLRYDLRD	2211
Db	2221	GPYANTTKTYDDGQGLQSVAVNDPRTWRYSYDNLGNLHLLNPGNSVRLMPLRYDLRD	2280
QY	2212	RITRLGDVOYKIDDDGYLCQSGDIFEYNSKGLLTRYANKASGWSYQYRDYGVGRASYK	2271
Db	2281	RITRLGDVYKIDDDGYLCQSGDIFEYNSKGLLTRYANKANGVYQYRDYDGLGRASCK	2340
QY	2272	TNLGHLHQLFYSDLNHPTRITHVYHNSSEITSLYDLOGLHLPAMSSGSEYVYVADSNK	2331
Db	2341	TNLGHLHQLFYADLNHPTRITHVYHNSSEITSLYDLOGLHLPAMSSGSEYVYVADSNK	2400
QY	2332	GTPPLAVFSLNGLMIKOLQYATGEIYYDSDNPQFQWIGVHGGGLYDPLTKLVHFTQRDYD	2391
Db	2401	GTPPLAVFSLNGLMIKOLQYATGEIYYDSDNPQFQWIGVHGGGLYDPLTKLVHFTQRDYD	2460
QY	2392	LGRWTSPTYTMKMYGKEPAPENLYMFKSNPNLSSELDELKNYVTDVKSWMVFGFQLSN	2451
Db	2461	LGRWTSPTYTMKMYGKEPAPENLYMFKSNPNLSNELDLKNYVTDVKSWMVFGFQLSN	2520
QY	2452	IIPGFPRAKMYFVPPYELSEQASENGQITGVQOTTERRHQAFMALEGQVITKKLHAS	2511
Db	2521	IIPGFPRAKMYFVPPYELSEQASENGQITGVQOTTERRHQAFMALEGQVITKKLHAS	2580
QY	2512	IREKAGHWFATPTPIIGKGMFAIKGGRVTTGVSSIASDSDSRKVASVLNNAYYLDKMHYS	2571
Db	2581	IREKAGHWFATPTPIIGKGMFAIKGGRVTTGVSSIASDSDSRKIASVLSNAHYLEKMHYS	2640
QY	2572	IEGKTHYFVKIGSADGDLVTLGTTIGRKVLESQVNVTVSQPTLLVNGRTRFTNIEFOY	2631
Db	2641	IEGKTHYFVKIGSADGDLVTLGTTIGRKVLESQVNVTVSQPTLLVNGRTRFTNIEFOY	2700
QY	2632	STLLLSIRYGLPTDLDDEKARVLDQARALCTAWAKEQQKARDGREGSRWLWTEGEKQ	2691
Db	2701	STLLLSIRYGLPTDLDDEKARVLDQARALCTAWAKEQQKARDGREGSRWLWTEGEKQ	2760
QY	2692	LLSTGRVQVGEYVLPVPEQYPELADSSNIQFLRONEMGRK 2733	
Db	2761	LLSTGRVQVGEYVLPVPEQYPELADSSNIQFLRONEMGRK 2802	
RESULT	4		
Q9WTS6			
ID	Q9WTS6	PRELIMINARY;	PRT: 2715 AA.
AC	Q9WTS6;		
DT	01-NOV-1999 (TremBLrel. 12, Created)		
DT	01-NOV-1999 (TremBLrel. 12, Last sequence update)		
DT	01-DEC-2001 (TremBLrel. 19, Last annotation update)		
DE	TEN-M3.		
GN	OD23 OR TEN-M3.		
OS	Mus musculus (Mouse).		

OC	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	
OX	NCBI_TaxID=10090;	
RN	(1)	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=BALB/C; TISSUE=BRAIN;	
RA	Ohashi T., Zhou X., Feng K., Richter B., Moergelin M., Perez M.T.,	
RA	Su W., Chiquet-Ehrismann R., Rauch U., Faessler R.;	
RT	*Mouse Ten-m/Odz is a new family of dimeric type II transmembrane	
RT	proteins expressed in many tissues.*;	
RL	J. Cell Biol. 0:0-0(1999).	
DR	EMBL; AB025412; BAA77398.1; -.	
DR	HSSP; P01492; IVNB.	
DR	MGI; MGI:1345183; Odz3.	
DR	InterPro; IPR000561; EGF-like.	
DR	InterPro; IPR002049; Laminin_EGF.	
DR	Pfam; PF00008; EGF; 6.	
DR	PRINTS; PR00011; EGF_LAMININ.	
DR	SMART; SM00181; EGF; 6.	
DR	PROSITE; PS00022; EGF_1; UNKNOWN_8.	
DR	PROSITE; PS01186; EGF_2; 7.	
KW	EGF-like domain; Glycoprotein.	
SQ	SEQUENCE 2715 AA; 303062 MW; 598F46A77334C2E1 CRC64;	
Query Match	70.7%; Score 10309; DB 11; Length 2715;	
Best Local Similarity	67.4%; Pred. No. 0;	
Matches 1883; Conservative 378; Mismatches 395; Indels 136; Gaps 16;		
QY	1	MDVKRR-HRSLTRGCGKECRYTSSSLSDSEDCRVPTQKSYSSSETLKAYDHD-SRMHYG 58
Db	1	MDVKRRPYCSLTSSRREKERRYTNSSADNEECRVPTQKSYSSSETLKAFDHYSLLYG 60
QY	59	NRVTDLIHRESDEFPQGTNFTLAEIGICEPSPHRS-GYCSDMGILHOGYSLSTGSDADS 117
Db	61	NRVKDLVHREADEYTRQGNFTLRQLGVCESTRRGVAFCAEMGLPHRGYSISAGSDAT 120
QY	118	DTEGGSPEHAIILMGRIKSRSSGLSRENSALTLTDSNENKSDDENGRPIPTSSP 177
Db	121	ENEAIVSPEHAMRLMGVKGSRSSCLSSRSNSALTLTDEHENRSDSEQP----- 173
QY	178	SLLPSAQLPSSINPPVPSQMPLLDSNTSHQIMDTNPDEEFPSPNSVLLLRACSGPQOASS 237
Db	174	-----SNP----- 177
QY	238	GPNHHSQSLRPLPPPHNHTLSHHSSANSLNRRSLNRRSQIHA-PAPAPNDLATTTP 296
Db	178	-----GQPTLQ-PLPPSHKQHPAQHPHSITSLNRRSLNRRNQSPAPPAALPAELQTTTP 230
QY	297	ESVQLQDSVWLSNVNPLETRHFLFKTSSGSTPLFSSSSPGYPLTSGTYVTPPRLLPRT 356
Db	231	ESVQLQDSVWLSNVNPLESRHFLFKTGTTPLFSTATPGYTMASGSVSPPTPLPRT 290
QY	357	FSRKAFKLKPKSKYKCAALSAATAALLLAILLAYETI-----VPWSL----- 400
Db	291	LSRSAPFKKSKYKSWRCTLCAVGVSVLLAILLSYFIAMHLFGLNHLQOTENDTFEN 350
QY	401	-----KNSIDSGEAEVGRVVTQEVPPGVFWSQIHI 432
Db	351	GKVNSTVPTNTVSLPSDNGKLGFTHENNTIDSGELDIGRAIQEVPPGIFWSQLFI 410
QY	433	SQPFLLKFNISLGKDALFGVYIRRGPLPPSHAQYDFMERLDG-----KEKWSVESPRERR 487
Db	411	DQPFLLKFNISLQKDALIGVYGRKGLPPSHQYDFVELLDGSRLLAREQNLVESERAGR 470
QY	488	STQTLVONEAVEVOYLDVGLVHLAFYNDGKDKEMYSFNTVLDSDQDCRNCHNGECYS 547
Db	471	QARSVSLHEAGFIQYLDSGIWHLAFYNDGKNPEQSFNTIVIESVVEECRNCHNGECYS 530
QY	548	GYCHCFPGFLGADCAKACPVLCSNGQYSKGTCCQYSGWKGAECDVPMNQCIDPSCGCH 607
Db	531	GTCHECFPGFLGDCSRACPVLCSSNGQYSKGRCLCFSGWKGTECDVPTTCIDPCGGR 590

QY 608 GSCIDGNCVCSAGYKGBCEBEVDCLDPTCCSHGVGVNCECLCSPGGLNCELARVOCPD 667  
Db 591 GICIMGSCACNSGVKGBCEBEADCLDPCGCSNHGVCIHGECCHNPGMGWCSNCEILTKWCAD 650  
QY 668 QCSGHGTYLPDPTGLCSDPNWMPGDCSVESVDCGTHGVICIGACRCBGWTAACDQR 727  
Db 651 QCSGHGTYLQESGCTCDPMTGDCSNEICSDVDCSHGCMGSCRCBGWTPACNQR 710  
QY 728 VCHPRCTEHCCTCKDKCECREGNGEHCCTIGROTAGTETDCCPLDCLNGRCRTLQNSWQ 787  
Db 711 ACHPRCAEHCTCKDKCECSOGNGEHCCTAHYLDKIVKEGCPGLCNSGRCRTLDQNGWH 770  
QY 788 CYCOTGHRGPGCNVAMETSCADKNDEGDLVDCLDPCCLQACQNSLLCRSRDPLDI 847  
Db 771 CVCPQWRGAGCDVAMETLCTDSKDNEDGLIDCMDDPCCLQSSCQNPYCRGLPDPDI 830  
QY 848 IQQG-QT-DHPAVKSFYDRILKLAGKDSHTHIGENPFNSLSVLIRGVVYTDGTPLVG 905  
831 ISOSLOTQPSOAAKSFYDRISFLGSDSTHVLPGESPFNSKSLASVIRGQVLTADGTPLLG 890  
QY 906 VNVSEFYKPYGYTITRQDGTFOLIANGASLTJLHFERAPFMSOERTVWLPWNSFYAMDT 965  
Db 891 VNVSEFLHYSEGYTITRQDGMEDLVANGASLTJLVEERSPFLQYHTVWLPWNVFYVMDT 950  
QY 966 LVNKTENSTPCDLSGFVRPDPITIISSPLSTFTSAAPOGNIPVPOVLHHEIELPGSN 1025  
Db 951 LVNKKENDIPSCDLSGFVRPSPITVSSPLSTFTSRSSPEDSPIPETQVLEHEETIPGTD 1010  
QY 1026 VKLYLSRRTAGYKSLKITMTOSTVPLNLTIRVHLMAVAGHLPKSFQSPNLAFTFIW 1085  
Db 1011 LKLSYLSRAAGYKSVUKITMTQAVIPFNLMKVHLMAVAVGRUFQKFPASPNAFYTFIW 1070  
QY 1086 DKTDAYQORYGLSDAVSVGYEYETPCSLILWEKRTALLQGFELDPNSLGGWSLDRKHI 1145  
Db 1071 DKTDAYNQYVGLSEAVSVGYEYESCLDLTWEKRTAVLQGYELDASNGGWTLDKHHV 1130  
QY 1146 LVNKSGLHKGTEGNOPLTOOPALITISIMGNGRRSISCPSCNGLAGENKLLAPVALAVG 1205  
Db 1131 LDVONGTLYKNGENOPISQOPPVYSSIMGNGRRSISCPSCNGAQDGNKLLAPVALACG 1190  
QY 1206 IDGSLYVGDENYTRIPSPSNVTSILERNKEPKHSNPAKHYYLAVDPVSGSLYVSDTN 1265  
Db 1191 IDGSLYVGDENYVNRIPSPSNVTSVLERNKNDFHSSNPAHRYLYLADPVTGDLVSDTN 1250  
QY 1266 SRRYRYSLSGDKLAGNESEVAGTGEQCLPFDEARCGDGGKAIDATLMSPRGIAVDKN 1325  
1251 TRRYRPKSLTGAKLTAKNAEVVAGTGEQCLPFDEARCGDGGRAVEATLMSPKGMAIDKN 1310  
1326 GLMYFVDATMIRKVDQNGIISTLLGSDNLTAVRPLSCDSSMDVAQVREWPTDLAVNPMD 1385  
Db 1311 GLIYFVDGTMIRKVDQNGIISTLLGSDNLTSAARPLTCDTSMHISQVREWPTDLAINPMD 1370  
QY 1386 NSLVLENNVILRTENHOVSIITAGRPMHCQVPGIDYLSLSKLATHSASATAISHTG 1445  
Db 1371 NSIYVLNNVVLQITENRQVRIAGRPMHCQVPGVEYIPVGHAVQITLESATAIAVSISG 1430  
QY 1446 VLYITETDEKKNIRLQVTTNGEITCLLAGAASDCCKNDVNCNCSGDDAYATDAIINSP 1505  
Db 1431 VLYITETDEKKNIRQVTTDGEISLVAGIPSECDCKNDANDCQYQSGDGYAKDAKLNP 1490  
QY 1506 SSLAVAPDGYIYIADLNIRIRAVSKNKPVLNAPNQYEAASPGQEILYVFNADGHIQYTV 1565  
Db 1491 SSLASPDGYIYIADLNIRIRAVSKNKPVLNAPNQYEAASPGQEILYVFNADGHIQYTV 1550  
QY 1566 SLVTGEYLYNFTYSTDNDVTELDNNGNSLKIIRDSSGMPRHLLMPDNOIITLTVGTNGG 1625  
Db 1551 SLVTGDLNLYNFSYSDNDVTAVTDNSGNTLRIRDPMRPMVVRVSPDNQVILWTIGTNGC 1610  
QY 1626 LKVYSTQNLEGLMGLTYDGNLTGLATKSDGTGWTTFYDHEGRLTNTVTRPTGVVYTSLHRE 1685  
Db 1611 LKSWTAGLEVLFTYHGNSSGLATKSDGTGWTTFYDSEGRLTNTVTRPTGVVYTNHGD 1670  
QY 1686 MEKSITIDIENSRDDVTITNLSSVEASYVYVQDVNRNSYQLCNNGTTLRVNVANGMI 1745

Db 1671 MDKAITVDIESSKREEDVSITSSLDSPYTWVQDLRNSYQIGYDGLSRIFVASCLDS 1730  
QY 1746 SFSEHSEHVLAGTITPTIGRCNISLPMENGLNISEWRLRKEQIKGKVITIFGRKLVRHGRNL 1805  
Db 1731 HYOTEPHVLVAGTANPTVAKRNMTPGCGENQNLVFERFKEQAGQKVNFKRKLVRNGRNL 1790  
QY 1806 LSTDYDNRINTEKIYDDHRRKFTLRIYDQVGRPFMLPSSGLAAVNVSYFNGRLAGLQR 1865  
Db 1791 LSVDFDTTKTEKIYDDHRRKFLRIAYDTSGHPTLMLPSSKLMANVYVYSTGOIASIOR 1850  
QY 1866 GAMSERTDIDKQIRIVSRMFADKVMWSYSYLDKSMVLLLSQSQRYIPEYDSSDRLLAVTM 1925  
Db 1851 GTTSEKVDYDSQGRIVSRVFEADGKTYSYLYLEKSMVLLLSQSQRYIPEYDMDRLSAITM 1910  
QY 1926 PSVARHSMSTHSIGYIRNLYNPESNASVIFDYSDGRILTKTSFLGTGRQVFKYKGLS 1985  
Db 1911 PSVARHTMQTIRSIGYIRNLYNPESNASIITDYNEEGLLLQTAFLGTSRRVLFKYRRQT 1970  
QY 1986 KLSEIYVDSTAVTFGVDETTGVLMKMNVLQSGFSCCTIRYRKIGPLVDKQIYRFSEEGMVN 2045  
Db 1971 RLSEILYDSTRVSFTYDETAGVLKTVNLQSDGFICTIRYRQIGPLIDRQIFRESEDCMVN 2030  
QY 2046 ARDYTYHDSNPRIASIKPVISETPLPVDLYRYDEISGKVEHFGKFGVYIYDINOITTA 2105  
Db 2031 AREFDYSY-DNSFRVTSMQGVINETPLPIDLYQFDDISGKVEQKFGVYIYDINOITTA 2089  
QY 2106 VMLTSKHEDTHGRKEVQYEMERSLVMYMTVOYDSMGRVVKRELKLGPAVNTKTYTYDYD 2165  
Db 2090 VMYTKHFDHGRKEVQYEMERSLVMYMTVOYDSMGRVVKRELKLGPAVNTKTYTYDYD 2149  
QY 2166 GDGQLQSVAVNDRTYRYSYDLNGLNLLNPGNSVRLMPLRYDLDRITRLGDVQYKIDD 2225  
Db 2150 VDGOLQTVYLNKIMTWRYNYDLNGLNLLNPPSSARLTPLRYDLDRITRLGDVQYRLDE 2209  
QY 2226 DGYLCQBGSDIPEYNSKGLLTRYNKNASGVSQYRYDCVGRARRSYKTNLGHHLQYFYSDL 2285  
Db 2210 DGYLRQGTETFEYSKGLLTRYNKNASGVSQYRYDCVGRARRSYKTNLGHHLQYFYSDL 2269  
QY 2286 HNPTRITHVYNSHSETSYLYDQHLFAMESSSGSEYVVASDNTGTPLAVFSINGLMI 2345  
Db 2270 TYPRTHVYNSHSETSYLYDQHLFAMESSSGSEYVVASDNTGTPLAVFSINGLMI 2329  
QY 2346 KOLQYTAIGEYIYDSDNPFQWVIGFHGGLYDPLTKLVHFTQRYDYVLAGRWTSPDYTMWK 2405  
Db 2330 KQOYTAIGEYIYDSDNPFQWVIGFHGGLYDPLTKLVHFTQRYDYVLAGRWTSPDYTMWK 2389  
QY 2406 NVGKEPAPPNLYKFSNNPLSSELDLKNYVTDVKSWSLVMFGFQLSNLIPIGPFRKMYFVP 2465  
Db 2390 RIGKDPAPPNLYKFSNNPLSSELDLKNYVTDVKSWSLVMFGFQLSNLIPIGPFRKMYFVP 2449  
QY 2466 PPELSESOASENGOLITGVQOTTERHNOAFMAL-EGOVITKKURASTREKAGHWA 2521  
Db 2450 PSELYKSOQWEDVPFFGVQOQVQVARAKAFSLGKMAEVQVSRKAGA---EGSWLWFA 2506  
QY 2522 TPTPIIGKIMFAIKRGTVTGVSSTIASDSRKVASVNNAYLDMKMHYSIEGKDTHYFV 2581  
Db 2507 TVKSLICKGVMLAVSQGRVQTVNLNANEDCIVAAVNNAFYLENLHFTIEGKDTHYFI 2566  
QY 2582 KIGSADGOLVLTGTTIGRKVLESQVNVTVSQPTLLVNGTRRRRTNIEFQVSTLLSRYG 2641  
Db 2567 KTTTPESDLGTLRLTSGRKALENGINVTVSQSTTVVNGTRRRFADVEMQFAGALAHRYG 2626  
QY 2642 LTPDPTLEEKARVLDOARQALGTAWAKEEQKARDREGSLRMTTEGKQOLLSTGRVQY 2701  
Db 2627 M---TLDEEKARILEQARALARAWAREQOQVRDREGGARLMTTEGKQOLLSTGRVQY 2683  
QY 2702 EGYVVLVPEQYPELADSSNIQFLRQNEGMGR 2733  
Db 2684 DGYVLSVEQYPELADSSNIQFLRQNEGMGR 2715

RESULT 5













Qy	1320	I	A	V	D	K	K	L	M	F	V	D	A	T	M	I	R	K	V	D	O	N	G	I	I	S	T	L	L	G	S	N	D	I	T	A	V	R	P	L	S	C	O	S	S	M	D	V	A	O	V	R	L	E	W	P	T	D	L	1379
Db	1179	I	A	V	D	K	K	F	I	F	V	D	G	T	M	I	R	K	V	D	R	N	G	I	I	S	T	L	L	G	S	N	D	I	T	S	A	R	P	L	T	C	O	N	S	M	I	G	O	V	R	L	E	W	P	T	D	L	1238	
Qy	1380	A	V	N	P	M	O	N	S	L	V	L	E	N	N	V	L	I	T	E	N	H	O	V	I	I	T	A	G	R	P	M	C	O	V	P	G	I	D	Y	S	L	S	K	L	A	I	S	A	E	S	A	I	1439						
Db	1239	A	I	N	P	M	O	N	S	I	V	L	O	N	N	V	L	O	T	E	N	R	O	V	I	V	I	A	G	R	P	M	C	O	V	P	G	I	E	T	M	G	K	R	A	I	O	T	T	L	E	G	A	T	1298					
Qy	1440	A	I	S	H	T	G	V	L	I	T	E	D	E	K	K	I	N	R	L	R	O	V	T	T	N	G	E	I	C	L	L	A	G	A	S	D	C	K	I	N	D	V	A	N	C	Y	S	G	D	A	Y	A	T	D	1499				
Db	1299	S	L	S	V	S	G	V	L	I	A	E	T	D	E	K	K	I	N	R	I	R	O	V	S	T	D	G	E	I	S	H	L	A	C	A	P	S	D	C	K	N	A	N	D	C	O	Y	T	G	D	G	Y	A	K	D	1358			
Qy	1500	A	I	N	S	P	S	S	L	A	V	A	D	G	T	I	I	A	D	L	G	N	T	I	R	A	V	S	K	N	K	P	V	L	A	N	F	N	O	Y	E	A	A	S	P	G	E	O	L	I	V	F	N	A	D	G	1559			
Db	1359	A	R	L	N	A	P	S	S	L	V	S	D	G	T	L	L	A	G	N	T	I	R	A	I	R	A	I	R	N	R	P	O	G	S	S	G	L	F	E	V	A	S	P	A	S	O	E	L	V	F	O	S	N	G	1418				
Qy	1560	I	H	O	V	T	S	L	V	T	G	E	V	L	N	F	T	S	T	O	N	D	T	E	L	D	N	G	S	L	K	I	R	R	O	S	S	G	M	P	R	H	L	L	M	P	O	I	T	L	T	1619								
Db	1419	T	H	O	T	W	S	L	V	T	G	Y	K	N	F	S	I	N	E	D	D	T	A	V	T	D	S	G	N	T	L	R	V	R	D	P	N	M	P	R	V	I	A	P	D	N	O	V	I	T	1478									
Qy	1620	V	G	T	N	G	L	K	V	V	S	T	O	N	L	E	L	G	M	T	Y	D	G	N	T	G	L	L	A	T	K	S	D	T	G	W	T	T	F	Y	D	H	E	G	R	L	T	N	T	R	P	T	G	V	1679					
Db	1479	I	G	T	N	G	L	K	T	L	A	O	Q	E	L	V	L	T	F	H	G	S	L	L	A	T	K	S	I	O	I	G	W	T	T	F	Y	D	S	E	G	R	L	T	N	T	F	T	G	V	1538									
Qy	1680	T	S	L	H	R	E	M	X	S	I	T	I	D	E	N	S	R	D	D	O	V	T	I	N	L	S	V	E	A	S	Y	T	V	O	D	O	V	R	N	S	V	O	L	C	N	G	T	L	V	M	Y	1739							
Db	1539	T	S	L	I	G	E	M	O	R	A	L	T	V	D	I	E	T	S	G	R	D	D	S	T	I	N	L	S	I	D	S	F	T	L	V	O	D	O	L	R	N	S	V	O	G	Y	D	N											

RESULT	Q9W7R3
ID	Q9W7R3
AC	Q9W7R3
OT	01-NOV
DT	01-NOV
DE	01-NOV
TE	TEN-MED
GN	TENN4
OS	Brachi
OC	Eukary
OC	Actinob
OC	Cyprinid
OX	NCHB-
RN	[1]
RP	SEQUENCE
EX	MEDIA
RA	Medea
RT	"Comp
RT	homolog
RT	systems
RL	Mech.
DR	EMBL;
DR	HSSP;
DR	ZFIN;
DR	Interf
DR	Interf
DR	Interf
DR	Pfam;
DR	Pfam;
DR	PRINTS
DR	SMART
DR	SMART
DR	PROSI
DR	PROSI
DR	EGF-1
SQ	SEQUENCE
KQ	



```
QY 2161 TYDYGDLQSVAVVNDRTWRYSYDLNGLHLLNFCNSVRLMLPLRYDLRDRITRLGDVQ 2220
Db 2248 AVEYDADGQLQVYSINDKPLWRYSYDLNGLHLLSPGNSARLTPLRYDIRDRITRLGDVQ 2307
QY 2221 YKIDDDGYLCQSGSDIFENYKGLLTRYANKASGVSQVRYDVGGRASYKTNLGHLLQY 2280
Db 2308 YRLDEGDLQRQGNDFEFYNSAGLLVKTYNKVNWMTIKYRIDGLGRVRSRSRQGHLLQF 2367
QY 2281 FYSDLNPTTRITHVYHNSSEITSLYDLOGLHFAMESSGGEYVYASDNTGTPLAVERSI 2340
Db 2368 FYADLSPTRVTHVYHNSSEITSLYDLOGLHFAFAMELSSGDEFFYACDNIGTPLAVESG 2427
QY 2341 NGLMIKQLQYATGEIYYDSNPDFQWVIGFHGGLYDPLTKLVHFTQDXYDVLAGRWTSPD 2400
Db 2428 AGLMIKQILHTAFGEYVLDSPFQVIGYQGGLYEPLTKLVHMGRRDYDVLAGRWTTPD 2487
QY 2401 YTMKWKNGKEP-APPNLYMFKSNPNLSSELDLKNYVTDVYKSWLVMFGFQLSNIIPGPRA 2459
Db 2488 HDIRKRLNSDNVYFNLVYMKNNPLNSQETKCYMTDVSMLVTFGFQLYNVIPGYRKP 2547
2460 KMYFVPPVPELSESOAS---ENGOLITGVQOOTTERHNOAFMALE--GOVITKK----- 2507
Db 2548 VTDAMEPSELVHTQIKTOEWDSKSVLGVCQEVQRLKSFVRLERFGQIYASDSGCPP 2607
QY 2508 --LHASIREKAGHWFATTPIIGKIGMFATKEGRVTTGVSSIASSEDSRKYVASVNNAYYL 2565
Db 2608 TPLHT-----LPATGTSLFGKGVKVAIREGRVEADIIISLANEDGRIIAVLDKASYL 2659
QY 2566 DKMHSIEGKDHIFYFKISADGLVTLGTTGRKVLSEGVNVTVSQPILLVNGRTRRT 2625
Db 2660 QDLHFTIAGLDTHIFYFKSLVEGLSLGTMVQGRLETGVNVTVSQVNVVLGGRSRRIT 2719
QY 2626 NTEFOYSTLLSTRYGLTPTDLDEEKARYLDOARALGTAWAKEOQKARDGREGSRLWT 2685
Db 2720 DIOMQVGTLSLVRYG---SSVDEEKVRVLELARQAVATAWAHERHRLRQEGESRAWT 2776
QY 2686 EGEKQOLLSTGRVQGYEGYVVLPEVQPELADSSSNIQFLRQNMGR 2733
Db 2777 DGERQOLLSSGRVQGYEGYVIVSDQFPELTIDNNVHEWQTEMGR 2824

RESULT 9
Q9JLC1 PRELIMINARY; PRT; 2346 AA.
AC Q9JLC1;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
OD23 (FRAGMENT).
OD23.
Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
SEQUENCE FROM N.A.
RA Ben-Zur T., Felge E., Motro B., Wides R.;
RP "The mammalian Odz gene family: Homologs of a Drosophila pair rule
RT gene with expression implying distinct yet overlapping developmental
RT roles.";
RL Dev. Biol. 117:107-120(2000).
DR EMBL; AF195418; AAF28316.1; -.
DR HSSP; P01492; LVNB.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002049; Laminin_EGF.
DR Pfam; PF00008; EGF; 6.
DR PRINTS; PR00011; EGF_LAMININ.
DR SMART; SM00181; EGF; 5.
DR PROSITE; PS00022; EGF_1; UNKNOWN_8.
DR PROSITE; PS01186; EGF_2; 7.
KW EGF-like domain; Glycoprotein.
FT NON_TER 1
```

```
SQ SEQUENCE 2346 AA; 261699 MW; D761563C6AEDD735 CRC64;

Query Match 63.2%; Score 9209; DB 11; Length 2346;
Best Local Similarity 70.7%; Pred. No. 0;
Matches 1657; Conservative 327; Mismatches 326; Indels 34; Gaps 9;

QY 401 KNSIDSAGEAEGRRVTQEVPGVFWRSQTHISQPOFLAKFNISLKGKDALFGVYIRGLPP 460
Db 26 ENNTIDSGELDIGRAIQEVPGIFWRSOLFIDPOFLAKFNISLQKDALIGVYGRGLPP 85
QY 461 SHAOYDFMERLQD----KEKNSVWSPRRRSIQTLVQNEAVFYQVLDVGLMHLAFYND 515
Db 86 SHTOYDEVLLDGSRLIAREORNLVESERAGROARSVLSHEAGFIQYLDLSGIWHLAFYND 145
QY 516 GKDKEMSVNTVVLDSVQDCPRNCHNGECVSCVCHCFGFLGADCAKACAPVLCSGNQ 575
Db 146 GKMPGVSENTIVIESVVECPRNCHNGECVSGTCHCFGFLGPDGCSRAACPVLCSGNQ 205
QY 576 YSKGTCCQYSGWKAGBCDVPMNQCIDPSGGHGCSDGNCVCAGYKGBHCEBVDCLDPT 635
Db 205 YSKGRCLCFSGWKGTEDVPTTQCIDPQCGRGICIGSCACNSGYKGENCEADCLDPC 265
QY 636 CSHGVCVNGECLSPGWGGLNCELARVQCPQDCSGHGYLPTDGLCSDPNMNGPDCSV 695
Db 266 CSHHGVCIHGECHCNPGWGSNCEILKTCADQCSGHGYLQESGSCDTPNMTGDCSN 325
QY 696 EYCSVDCGTHGVICIGACRCEEGWTCAACDORVCHPCIEHGTCCKGCKCECEGNGEHC 755
Db 326 EYCSVDCGSHGVCMGSCRCEEGWTGACNORACHPCAEHGTCKGCKCECEGNGEHC 385
QY 756 TIGROTAGTETDGCPLDCNGNRCITLQNSWCVCQCTGMRGPGCNVAMETSCADNKDNEG 815
Db 386 TI-----EGCPGLCNSNGRCTLDQNGHVCVCPQWGRGACGCDVAMETLCTDSKDNEG 436
QY 816 DGLVDCLDPPCLQSAQNSLLCRGSRDPLDIIQQG-QT-DWPAVKSFFDRIKLLAGKDS 873
Db 437 DGLIDCMDPPCLQSSCQNPYCRGLPDPQDIIISQSLQTPSQQAASKFYDRISFLSGDS 496
QY 874 THIPGNPNSLSLIRGVVTTDGTPLVGVNVFVKYPKYGYTTTQDGTGFDLIANG 933
Db 497 THVLPGESPFNKSLASVIRGOVLTDGTPPLGVNVFSLHYSEGYTTTQDGTGFDLVANG 556
QY 934 GASLTLHFERAPMSQERTVWLPWNSFYAMDTLVMKTEENSPISCDLSGFVRPDPPIISS 993
Db 557 GASLTLVFRSPPLTQYHTWIPWNVFYMDTLVMKKEENDIPSCDLSGFVRPSPIISS 616
QY 994 PLSTFFSAAPGQNPVTPETQVHLHEIELPGSNVVKLRYLSRRTAGYKSLKITMTQSTVPL 1053
Db 617 PLSTFFSRSPEDSPIIPTQVHLHEETTIPGTDUKLSYLSRAAGYKSVLKITMTQAVIPF 676
QY 1054 NLIRVHLMAVEGHLFQKSFQASPNLASTFIMDKTDAYGORVYGLSDAVYVSGFEYETCP 1113
Db 677 NLMKVHLMAVAVYGRLEFQKWFPPASPNLAYTFIMDKTDAYNQYVGLSEAVYVSGFEYESCL 736
QY 1114 SLTLWEKRTALQGFELDPNSLGGWSLDKXHHILNVKSGILHKGTGENQFELTQPAITSI 1173
Db 737 DLTLEWKRTAVLQGYELDASNMGMGWTLDKHHVLDVQNGILYKNGENQFISQPPVYSSI 796
QY 1174 MNGRRRSISCPSCNGLAEGNKLAPVALAVGIDGSLYVCDENYIRIIPSRNVTSILEL 1233
Db 797 MNGRRRSISCPSCNGQADGNKLLAPVALACGIDGSLYVCDENYIRIIPSPGNVTSVLEL 856
QY 1234 RNKEFKHSNNPAHYLYLAVDPVSGSLYVSDTNSRRIRYRKSLSGTKDLAGNSVWAGTGE 1293
Db 857 -----SSNPAHYLYLATDPTVTDGLYVSDTNRRIYRKSLTGAKDLTKNAEVVAGTGE 909
QY 1294 QCLPFDEARCGGKKAIDATLMSPRGIADVKNGMLTFVDMTRKVDQNGIISTLGSND 1353
Db 910 QCLPFDEARCGGKKAVEATLMSPKGMAIDKNGLIYFVDGTMIKRVQDQNGIISTLGSND 969
QY 1354 LTAVRPLSCDSSMDVAQVLEWPTDLAVPMDSLVLENVILRLITENHOVSIIAGRPM 1413
Db 1413
```

```
Db 970 LTSARPLTCDTSMHISQVRLEWPTDLAINPMDNSIYYLDNNVVLQITENRQVRIAAGRPM 1029
Qy 1414 HCQVPGIDYSLSKAIHSALESASAIASHGTGLVLIYITETDEKKINRLRQVTTNGEICLLA 1473
Db 1030 HCQVPGVEYVPGKHAVQTTLESATAIAVSYSGVLVIYITETDEKKINRLRQVTTDGEISIVA 1089
Qy 1474 GAASDCCKNDVNCNYSGGDAYATDAILNSPSSSLAVAPDGTIYIADLGNIRIRAVSKNK 1533
Db 1090 GIPSECDCKNDANGDCYQSGDGYAKDARLAPSSSLAASPDCGLTYIADLGNIRIRAVSKNK 1149
Qy 1534 PVLNAFNOYEASPGEOELYYFNADGTHOYTVSLVGTGEYLYNFTYSDNDVDTELIDNGN 1593
Db 1150 PLLNSMFYVASPTDQELIYFDLNGHTQYTVSLVGTGYLYNFSYNDNDVTAVTDNGN 1209
Qy 1594 SLKIRRDSSGMRHLLPNDQIITLTGTNGGLKXVWSTQNLEGLMTYDNGTGLLATKSD 1653
Db 1210 TLRIRDPNRPVVRVSPDNQVILWTGTNGCLAKSMTAQLGLVLFYHGSGLLATKSD 1269
Db 1654 ETGWTTTFDYDHEGRLLNVTRPTGVVTSLSHREMEKSITIDIENSNRDDDDVTVITNLSVE 1713
Db 1270 ETGWTTTFDYDSEGLTNVTPPTGVVTLNHLGDMDKAITVDIESSSREEDVSIITNLSID 1329
Qy 1714 ASYTVVODVNSYQLCNNGTFLRVYANGMGISPHSEPHVLAGNITPTIGRCNLSLPEN 1773
Db 1330 SFYTWVQDLNNSYQIGVDSGLRIFYASGLDSHTQTEPHVLAGTANPTVAKRNMTPGEN 1389
Qy 1774 GLNSIELRLRKEQIKGVITFGRLKRVHGRNLLSIDYDRNIRTEKIYDDHKKFTLRIYD 1833
Db 1390 GONLVENRFRKEQAOGKVNVEGRKLVNGRNLLSVDFDRTTKTEKIYDDHKKFTLRIAYD 1449
Qy 1834 QVGRPFVLWSSGLAANVSYFFNCRLAGLQAGMSEKTDKQGRIVSRMFADGKVMYS 1893
Db 1450 TSGHPTLWLPSSKLMANVTYSSTGQIASIQRTSEKVDYDSQGRIVSRVFADGKTWSY 1509
Qy 1894 SYLDKSMVLLQSORQYIFEDYSDRLAVTMPSPVARHSMSTHPSIGVIRNIYNPPESNA 1953
Db 1510 TYLEKSMVLLHRSORQYIFEDYDMDRLSAITMPSVARHTWOTIRSIGVIRNIYNPPESNA 1569
Qy 1954 SVIFDYSDGGRILKTSFLGTGRQVIFYKGLSKLSEIYDSTAVTFGDETGYLVKMWNL 2013
Db 1570 SIITDYNEEGLLQATFLGTSRRVLEKRYRQTRLSLEIYDSTRVSTYDETAVGLKTVNL 1629
Qy 2014 QSGFSCITIRYRKGLPLVDKQIYRFSSEGWNARFDYTHDNSEFRIASIKPVISETPLVY 2073
Db 1630 QSDGFICTIRYRQGLPLDRIQIFRSEDMGNARFDYSY-DNSFRVTSMQGVINETPLPI 1688
Qy 2074 DLYRYDEISGRVGHFGVGIYIDINQIITAVMTLSKHFDTHGRKEVQYEMERSLWY 2133
Db 1689 DLIQFDLISGRVGHFGVGIYIDINQIISTAVMTYTKHFDHGRKEIYEIERSLWY 1748
Qy 2134 MTVOYDSMGRVIRKRELKGPYANTTKYTYDYDGDGQLQSVAVNDRPTWRYSYDLNGLNHL 2193
Db 1749 ITIOYDNMGRTYKREIKIGPANTTKYAYEYDVGQLQTVYLNKIMWRYNDLNGNLHL 1808
Qy 2194 LNPNSVRLMPLRYDLDRITRLGDVQYKIDDDGYLCQSGDIEFYNSKGLLTRAYNKAS 2253
Db 1809 LNPSSARLTPLRYDLDRITRLGDVQYRLDDEGFLRQGTETFEYSSKGLLTRVYSKGS 1868
Qy 2254 GWSVOYRYDVGRRASYKTNLGHHLQVYSDLHNPTRTHVYNNHNSSEITSLYDLOQHL 2313
Db 1869 GWTVIYRYDGLGRRVSSKTSUGHLQFFYADLTPTPRTHVYNNHNSSEITSLYDLOQHL 1928
Qy 2314 FAMESSSGEYVASDNTGTPLAVFSINGLMIKQLQYTAYGEIYDSDPDMFQWIGFHHG 2373
Db 1929 FAMELSSGDEYIASDNTGTPLAVFSSNGLMKQIYQYAYGEIYFDSNVDFOLVIGFHHG 1988
Qy 2374 LYDPLTKLVHFTQDYDVLAGRWTSPDYTMWKNVKGCPAPFNLYMFKSNPNPLSSDELDKN 2433
Db 1989 LYDPLTKLVHFTGERDYDILAGRWTTPOTEIKWIKGDKPAPFNLYMFRNNPNASKIHDVKD 2048
Qy 2434 YTVDVKSHLVNMGFQLSNIIIGFPPRAKMYFVPPPYELSESQASENGLITGVQQTTERHN 2493
Db 2049 YITDVNSLVLTGFFHLHNAIFGFPVPKFDLTPEPSYELVKSQOWEDVPPIFGVOQVQARQA 2108
```

```
Qy 2494 QAFMAL----EGQVITRKLHASIREKAGHWFAATPTPIIGKIMFAIKRGRVTTGVSSIAS 2549
Db 2109 KAFSLGKMAEVQVSRKAGA---EQSLWFATVKSLLIGKVMGLAVSOGRVQTVNLNAN 2165
Qy 2550 EDSKRVASLVNAYYLDKMHYSIEGKTHYFVKTGSADGDLVTLGTTIGRKVLSSGVNVT 2609
Db 2166 EDCIKVAALNNAPYLENLHFTIEGKTHYFVKTTTPESDLGTLRLTSGRKALENGINVT 2225
Qy 2610 VSQPTLLVNGRTREFTNIEFOYSTLLLSIRYGLTPDTLDEEKARVLDOARORALGTAWAK 2669
Db 2226 VSQTTTVVNGRTTRFADVEMQFGALAHVRYGM---TLDEEKARLEQARORALARAWAR 2282
Qy 2670 EQQKARREGSRLEWTEGEKQQLLSTGRVQGYEGYVYLPVQYPELADSSNIQFLRONE 2729
Db 2283 EQQRVDGEGEARLWTEGEKQQLLSACKVQGYDCYVYVLSVQYPELADSSANNIQFLROSE 2342
Qy 2730 MGKR 2733
Db 2343 IGKR 2346

RESULT 10
Q9ULU2 ID Q9ULU2 PRELIMINARY; PRT; 1737 AA.
AC Q9ULU2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE KIAA1127 PROTEIN (FRAGMENT).
GN KIAA1127.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20039618; PubMed=10574461;
RA Hirosewa M., Nagase T., Ishikawa K., Kikuno R., Nomura N., Ohara O.;
RT "Characterization of cDNA clones selected by the Genemark analysis
from size-fractionated cDNA libraries from human brain.";
RL DNA Res. 6:329-336(1999).
DR EMBL; AB032953; BAA86441.1; -.
DR InterPro; IPR001258; NHL.
DR Pfam; PF01436; NHL; 2.
FT NON_TER 1
SQ SEQUENCE 1737 AA; 195065 MW; CFF0AF0C8C5D6B32 CRC64;
```

```
Query Match 62.1%; Score 9055; DB 4; Length 1737;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1736; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 997 TFFSAAPQNPVETQVLEHIEELPGSNVLRVLSSTAGYKSLKTKTMTQSTVPLNLI 1056
Db 1 TFFSAAPQNPVETQVLEHIEELPGSNVLRVLSSTAGYKSLKTKTMTQSTVPLNLI 60

Qy 1057 RVHLMVAVEGHFLFKSQAQSNLASTFTWDKTDAYGQRYGLSDAVSVSGFEYETCPSLI 1116
Db 61 RVHLMVAVEGHFLFKSQAQSNLAYFTWDKTDAYGQRYGLSDAVSVSGFEYETCPSLI 120

Qy 1117 LWKRTALLQCFELDPNSNLGWSLDKHHILNVKSGILHKGTTGENOFLTQQPAITISIGN 1176
Db 121 LWKRTALLQCFELDPNSNLGWSLDKHHILNVKSGILHKGTTGENOFLTQQPAITISIGN 180

Qy 1177 GRRSISPCSCNGLAEGNKLAPVALAVGIDGSLYVGDFNYIRIRFPSSNNVTSILELRNK 1236
Db 181 GRRSISPCSCNGLAEGNKLAPVALAVGIDGSLYVGDFNYIRIRFPSSNNVTSILELRNK 240

Qy 1237 EFKHSNNPAHKYIYLAQVPSGSLVSDTNSRRIRVKSLSGTDKLAGNSEVAVAGTGEOL 1296
Db 241 EFKHSNNPAHKYIYLAQVPSGSLVSDTNSRRIRVKSLSGTDKLAGNSEVAVAGTGEOL 300
```



QY 1297 PFDEARCGDGGRAIDATLMSPRGIAVDKNGLMYFVDATMIRKVDQNGIISTLLGNSDLTA 1356  
DB 301 PFDEARCGDGGRAIDATLMSPRGIAVDKNGLMYFVDATMIRKVDQNGIISTLLGNSDLTA 360  
QY 1357 VRPLSCDSDMDVAQVRLNPTDLAVNPMDSLVLENVILRITENHQVSIITAGRPMHCO 1416  
DB 361 VRPLSCDSDMDVAQVRLNPTDLAVNPMDSLVLENVILRITENHQVSIITAGRPMHCO 420  
QY 1417 VPGIDYLSLKLHSALESASAIATSHCTGVLYITETDEKKNRRLQVTTNGEICLLAGAA 1476  
DB 421 VPGIDYLSLKLHSALESASAIATSHCTGVLYITETDEKKNRRLQVTTNGEICLLAGAA 480  
QY 1477 SDCCKNDVNCYSGDDAYATDAILNSPSSLAVAPDGTIYIADLGNIRAIRVSKNKPVL 1536  
DB 481 SDCCKNDVNCYSGDDAYATDAILNSPSSLAVAPDGTIYIADLGNIRAIRVSKNKPVL 540  
QY 1537 NAFNQYEAASPGQEELYFVFNADGIIHQYTVSLVTGEYLYNFTYSTDNVTELIDNNGNSUK 1596  
DB 541 NAFNQYEAASPGQEELYFVFNADGIIHQYTVSLVTGEYLYNFTYSTDNVTELIDNNGNSUK 600  
QY 1597 IRRDSSGMPRHLLMPDQNIITLTVTGNGGLKVYSTQNLGLMTYDGNLGLATKSDETG 1656  
DB 601 IRRDSSGMPRHLLMPDQNIITLTVTGNGGLKVYSTQNLGLMTYDGNLGLATKSDETG 660  
QY 1657 WTTFFDYDHEGRLTNVRTPTGVVTSLSHREMEKSIITIDIENSRRDDVVTITNLSSVEASY 1716  
DB 661 WTTFFDYDHEGRLTNVRTPTGVVTSLSHREMEKSIITIDIENSRRDDVVTITNLSSVEASY 720  
QY 1717 TVVQDVNRSYOLCNGNGLTVRVYANGMGISFSEPHVLVAGTITPTIGRCNISLPHENGUN 1776  
DB 721 TVVQDVNRSYOLCNGNGLTVRVYANGMGISFSEPHVLVAGTITPTIGRCNISLPHENGUN 780  
QY 1777 SIEWRLRKEQIKGKVTIFGRKLVRHGRNLLSIDYDRNIRTEKIYDDHRKFTIRIYDQVG 1836  
DB 781 SIEWRLRKEQIKGKVTIFGRKLVRHGRNLLSIDYDRNIRTEKIYDDHRKFTIRIYDQVG 840  
QY 1837 RFPLMLPSSGLAAVNSVYFENGRLAGLQRGAMSERDIDKQGRIVSRMPADCKVMSYSYL 1896  
DB 841 RFPLMLPSSGLAAVNSVYFENGRLAGLQRGAMSERDIDKQGRIVSRMPADCKVMSYSYL 900  
QY 1897 DKSMVLLLOSQRYIFEXYDSSORLLAVTMPVSARVMSMTHSTIGYIRNIYNPESNASVI 1956  
DB 901 DKSMVLLLOSQRYIFEXYDSSORLLAVTMPVSARVMSMTHSTIGYIRNIYNPESNASVI 960  
QY 1957 FQYSDGRILKTSFELGTGRQVFKYKGLSKLSEIYVDSTAVTFGYDETTGVLMKNLQSG 2016  
DB 961 FQYSDGRILKTSFELGTGRQVFKYKGLSKLSEIYVDSTAVTFGYDETTGVLMKNLQSG 1020  
QY 2017 GFSCITIRYRKIGPLVDKQIYRSEEGVNNARFDYTYHONSFRIASIKPVISETPLPVDLY 2076  
DB 1021 GFSCITIRYRKIGPLVDKQIYRSEEGVNNARFDYTYHONSFRIASIKPVISETPLPVDLY 1080  
QY 2077 RYDEISGKVEHFGKGVIIYDINQIITAVMTLSKHFDTHGRIKEVQVEMFSLMYWMTV 2136  
DB 1081 RYDEISGKVEHFGKGVIIYDINQIITAVMTLSKHFDTHGRIKEVQVEMFSLMYWMTV 1140  
QY 2137 QYDSMGRVIRKELKLGPAVNTTKYTYDYDGDQGLQSVAVNDRPTWRYSYDLNGLHLLNP 2196  
DB 1141 QYDSMGRVIRKELKLGPAVNTTKYTYDYDGDQGLQSVAVNDRPTWRYSYDLNGLHLLNP 1200  
QY 2197 GNSVRLMPLRYDLRITRLGDOVQYKIDDDGYLCORGSDIPEYNSKGLLTRYNKAQSGWS 2256  
DB 1201 GNSVRLMPLRYDLRITRLGDOVQYKIDDDGYLCORGSDIPEYNSKGLLTRYNKAQSGWS 1260  
QY 2257 VOYRVDGVRGRRASYKTNLGHHLQYFYSDLHNPTRTHVYNHNSSETSLYDYLQGHLLFAM 2316  
DB 1261 VOYRVDGVRGRRASYKTNLGHHLQYFYSDLHNPTRTHVYNHNSSETSLYDYLQGHLLFAM 1320  
QY 2317 ESSSGEYVVASDNTGTPLAVFSINGLMIKQLQYATAYGEIYYDSNPDMQVIFGFGGLYD 2376  
DB 1321 ESSSGEYVVASDNTGTPLAVFSINGLMIKQLQYATAYGEIYYDSNPDMQVIFGFGGLYD 1380

QY 2377 PLTKLVHFTQRDYDVLGRWTSPTDYTMKNVKGKEPAPFNLYMFKSNPNLSSELDLKNYVT 2436  
DB 1381 PLTKLVHFTQRDYDVLGRWTSPTDYTMKNVKGKEPAPFNLYMFKSNPNLSSELDLKNYVT 1440  
QY 2437 DVKSWLVMEFGFQLSNIPGFPRAKMYFVPPPYELSSQASENGQLITGVQOQTTERRNQAF 2496  
DB 1441 DVKSWLVMEFGFQLSNIPGFPRAKMYFVPPPYELSSQASENGQLITGVQOQTTERRNQAF 1500  
QY 2497 MALEGOVITTKLHASIREKAGHWFATTTPLICKGIMFAIKREGRTVTGVSIIASEDSRKVA 2556  
DB 1501 MALEGOVITTKLHASIREKAGHWFATTTPLICKGIMFAIKREGRTVTGVSIIASEDSRKVA 1560  
QY 2557 SVLNNAVYLDKMHYSTEGKDTYFVKIGSADGDLVTLGTTIGRKVLESVNVTVSPTLL 2616  
DB 1561 SVLNNAVYLDKMHYSTEGKDTYFVKIGSADGDLVTLGTTIGRKVLESVNVTVSPTLL 1620  
QY 2617 VNGRTRFRNIEFOYSTLLSIRYGLTPTDDEKARVLDQARQALGTAWAKEQOKARD 2676  
DB 1621 VNGRTRFRNIEFOYSTLLSIRYGLTPTDDEKARVLDQARQALGTAWAKEQOKARD 1680  
QY 2677 GREGSRLMTEGEGKQOLLSTGRVQGYGYVLPVEQYPELADSSSNIQFLRQNMGR 2733  
DB 1681 GREGSRLMTEGEGKQOLLSTGRVQGYGYVLPVEQYPELADSSSNIQFLRQNMGR 1737  
RESULT 11  
Q9W6V6 PRELIMINARY; PRT; 2705 AA.  
AC Q9W6V6  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE TENURIN-1.  
GN TEN-1.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRIN, RETINA;  
RX MEDLINE=99276585; PubMed=10341219;  
RA Minet A.D., Rubin B.P., Tucker R.P., Baumgartner S.,  
Chiquet-Ehrismann R.;  
RA "Teneurin-1, a vertebrate homologue of the Drosophila pair-rule gene  
ten-m, is a neuronal protein with a novel type of heparin-binding  
domain.";  
RL J. Cell Sci. 112:2019-2032(1999).  
DR EMBL; AJ238613; CAB43098.1; -.  
DR HSP; P00750; ITPG.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR002049; Laminin\_EGF.  
DR Pfam; PF00008; EGF; 5.  
DR PRINTS; PR00011; EGFLAMININ.  
DR SMART; SM00181; EGF; 6.  
DR SMART; SM00001; EGF-like; 1.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_8.  
DR PROSITE; PS01186; EGF\_2; 7.  
KW EGF-like domain; Glycoprotein.  
SQ SEQUENCE 2705 AA; 302385 MW; 230F03D1999037D2 CRC64;

Query Match 58.6%; Score 8541.5; DB 13; Length 2705;  
Best Local Similarity 55.9%; Pred. No. 0;  
Matches 1569; Conservative 471; Mismatches 590; Indels 175; Gaps 26;

QY 1 MDVKRRHRSRLTRGCGKECRVTSLSLSEDCRVTQKSYSSSETLKAYDHDPSRMYGNR 60  
DB 4 MDCKP--YQPLSKVHEVDLTVTSSEDESDGKQKQ--SYDSRETINYSQELRLNY--- 57  
QY 61 VTDLHRESDEPFGTNTFLAELGICPSPH--RSQYCSDM-GILHQYSLSTGSDADS 117

Db 58 -----NSQSRKRKNTDQSTQDMFCE--TPHILCSGYQTDLHGVSSEHSYPLEVGSVDYT 109  
Qy 118 DTEGMSPEHAIRLWGRIGIKRRSSGLSSRENSALTLDSDNEKNSDDBENGRIPTPTSSP 177  
Db 110 ETEGASPDHALRMMRGCMKEHSSCLSSRANSALS LTD DHERKSDGENDMPGSPHNQF 169  
Qy 178 SLLPSAQPLSHNNPVSQMPPLDLSNTSHQIMDNPDEFSNYSLLRACSGPQQAASS 237  
Db 170 TFRP---Lp-----ppp----- 178  
Qy 238 GPPNHSOSTLRPLPPPHNHTLSHHSSANSLSNRSLTNRRSQIHAPAP-APNDLATTP 296  
Db 179 -PPHACTCTKRP-----PAASLQRRSMT-RSQ---PSPAAPTPTSTQ 220  
Qy 297 ESOVLQDSWVLSNVPLETRHFLFKTSSGSTPLFSSSPGYPLTSGTYVTPPPRLLPNT 356  
Db 221 DSVHLHNSWVLSNIPLETRHFLFKHSGSSAIFSAASONYPLTSNTVSPPPRLPNT 280  
Qy 357 FSRKAFKKKSKYCSWKAALSAIAAALLAILLAYI-----VPSWL----- 400  
Db 281 FSRPAFTSKPYRCNWKCTALSATAITVTLALLAYVIAVHLFGLTWQLOPVEGQLYEN 340  
Qy 401 -----KNSIDSGEAEVGRVTOEVPGPVWR 427  
Db 341 GVSCKNGKAESTDITYSPIGKVSVDTEKKVQKGRADTGEVEIGAQMOTIPPLGLEWR 400  
Qy 428 SQIHISOPQFLKFNISLQKDALFGVYIRRGPLPPSHAQYDFMERLDGKEKWSVSPRERR 487  
Db 401 FQITIHHPVYLFKNISLAKDSLGIYGRNIPPTHQTFDFVKLMDGKQ--LIKQEPKSE 458  
Qy 488 SIQTLVQN-----EAVFVQYLDVGLMHLAFYNDGKDKEWVSFNVVLDVSQDCPNRCH 540  
Db 459 EPQAPRNLILTSQETGFIEYMDQGAHMAFYNDGKVEQVFLVTTAIEVLDLDCSTNCN 518  
Qy 541 GNGECVSGVCHFCFGFLADCAKACAPVLCSGNGOYSGTKCQYSGWGAECDPVNMOCI 600  
Db 519 GNGECISGHCHFCFGFLPDCAKOSCVLCSGNGEYERGHCVCRNGWGPCECDVPEEOCI 578  
Qy 601 DPCSGHGSCIDGNCVCAGYKGEHCEBVDCLDPTCSHGVCVNGECILCSPGWGLNCEL 660  
Db 579 DPTCFGHGTCINGVICVPGYKGEICEEEDCLDPMC SGHVCVQGECHCSAGWGVNCT 638  
Qy 661 ARVOCPCDCSGHGYLPTDGLCSDDPNMGPDCSVVEGSDVCGTHGICGACRCEGWT 720  
Db 639 SLPICOEHCSGHGTFLDVLGCLSCRPQWGTGSDCSELCCTDCGSHGVCGRGICQCEGW 698  
Qy 721 GAACDORVCHPRCIBHGTCCKGCKBREGWNGEHCITIGRQTAGTETDCGCPDLCNGRGT 780  
Db 699 GPTCEERTCHSHAEBHQCKDGKCECSPWEGGDHCTI-----DCGPGLCYNGRGT 749  
Qy 781 LCQNSWOCVQCTGWRGPCNVMAMETSCADNKDNEGDGLVDCLDPCCLQSACQNSLLCRG 840  
Db 750 LDONGHVCQVQSGSCNVMVMACGNDLNDGDGLTDCVDPCCQONNCYASPLCQG 809  
Qy 841 SRDPLDIIQOQOTDWP--AVKSFYDRIKLLAKGDSHTHPIGNEPNNSSLVSLIRQVYTT 898  
Db 810 SPDPLDLIOHSPQSPHPRFLFYDRIFLIGKETHVIPGDISPESRRASVIRQGVVAI 869  
Qy 899 DGTPLGVNVSVFKYPKYGYTITRODGTFDLIANGCASLTLHFERAPFMSQERTVWLWPN 958  
Db 870 DGTPLGVNVSVFLHDEGYTISRODGSFDLVAVGGISVTLVDFDRSPFISEKRTLWLSWN 929  
Qy 959 SFYAMDTLVMTKEENSIPSCDLSGVRPDPPIIISPLSTFEESAAPQNPVETOVLEHE 1018  
Db 930 REVIVDKVMQRAESDIFSCDVSSPISPNPVLPSPLTAFGSCPERGTVIPLOVQOE 989  
Qy 1019 TELPGSNVRLYSSRTAGYKSLKITMTQSTVPVNLNLRVHLMAVEGHFLQKSFQASP 1078  
Db 990 IPIPSFVKLSVLSRSTPGYKTLRLVILHTIPTSGMFKVHLIIAVERGLLQKWFPAAN 1049  
Qy 1079 LASTIWDKTDAYGORVYGLSDAVSVGFEYETCPSLILWEKRTALLQGFELDPNSLGGW 1138  
Db 1050 LVYTFANWKTDIYGQKVSGLAEAMYSVGYEYETCPDFILWEKRTVILQGFENDASNLGGW 1109

Qy 1139 SLDKHHILNVAKSGILHKGTGENQFLTQOPAITTSIMNGRRRSISCPSCNGLAEGNKLLA 1198  
Db 1110 SINKHHVLNPOSGIVHKNGENMFISQOPPVISTMKGNGHQRSVSCSNGLANSKLFA 1169  
Qy 1199 PVALAVGDGSLYGVDFNYIRRIFFSRNWTISILELRNKEFKHNNPAHKYLYLAVDPVSGS 1258  
Db 1170 PVALTSGDPGSVYIGDFNEFVRRIFFPSGNSIGTLELRNDRTHSTSPAHKYLYLAVDPVSES 1229  
Qy 1259 LYVSDTNRRLYRVKLSLSTKDLAGNSEWAGTQOCPLPFEARGDGGKADATLMSPR 1318  
Db 1230 LYLSDTNTRRYKAKSLTETKDLAKNDVAVAGTQOCPLPFDQSHCGDGGKASELNSPR 1289  
Qy 1319 GIAYDKNGLMYFVDATMIRKVDQNGIISTLGLSNDLTAVRPLSCDSSMDVAQVRLEWPTD 1378  
Db 1290 GITIDKHGFIYFVDCGMIRKIDENGMITTIIGSNGLTSTQPLSCDSDGMDITQVRLEWPTD 1349  
Qy 1379 LAVPMDNSLYVLENNVILRITENHOVSIAGRPHMCOVPGID-YSLSKLAIHSALESAS 1437  
Db 1350 LTVNPLDNLNLYVDNNVLIQISERRVRIIAGRIHCOVPGIDHFIIVSKVAIHSTLESAR 1409  
Qy 1438 AIAISHTGVLYTETDEKKINRLROVTTNGEICLLAGAADCCDKNDVNCNCYSGDDAYA 1497  
Db 1410 AIAYSHSGIPIRETDERKINRIQOVTNGEISIIAGAPSDCCDKIDPNCDFSGDGYA 1469  
Qy 1498 TDAIINSPSSLAVAPDGTIYIADLGNIRIRAVSKNKPVLNAFNOYEAASPGEQELYVFN 1557  
Db 1470 KDAKLAPSSLAVSPDDTLYVADLGNIRIRAVSRNKAHLSDTNMYEIAIPADQELYQFTI 1529  
Qy 1558 DGIHQYTVSLVTGEVLYNFTYSTDNVTELIDNNGNSLUKIRRDSSGMPRHLLMPNQIIT 1617  
Db 1530 NGTHLHTNLNLTRODIYNTFYSGEDVATITSSNGNSVHIRRDTSGPLVWVVPQGVYW 1589  
Qy 1618 LTVCTNGGLKVYSTONLEGLWTDGNTGLLATKSDGTWTFYDYDHEGRLTNVTRPTG 1677  
Db 1590 LTISSNGVLKRYAOGYNLAWTYPGNTGLLATKSDENGTWTFYEDSGHLTNATFPTG 1649  
Qy 1678 VVTSLHREMEKSITDIENSRRDDVTITNLSSYEASVYVQDQVRNSYQLCNGTILRV 1737  
Db 1650 EVSFSHSDVEXLTVLELDTSNRENNVTA-TNFSATSTIYTLKQDNTQNIYRVSPDGLRV 1708  
Qy 1738 MYANGMGSFISEPHVHLAGTTPTTIGRGNISLPMENGLNSIEWRLURKQIKGVITFORK 1797  
Db 1709 TFASGMEITLATEPHILAGVSPITLKGKINISLPGHNSNLIEWRQREQTGNITFFERR 1768  
Qy 1798 LRVHGRNLLSDYDRNIRTEKIYDDHRAKFTLRIYDVGVRLPLPSSLAANVSYFFEN 1857  
Db 1769 LRAHNRNLLSIDFHVTRTGKIYDDHRAKFTURIMYDOTGRPVLWSPISKYNEVNITYSHS 1828  
Qy 1858 GRLAGLQRGAMSERDIDKQGRIVSRMFADQKVSYSYLDKSMVLLLSQSQRYIFEYDSS 1917  
Db 1829 GLVTVYIQRGTWTEKXEYDPSGNIISRTWADGKIWSYTYLEKSMVLLLSHQRRYIFEYDOS 1888  
Qy 1918 DRLAVTWPSVARHSMSTHTSIGYIRNYNPPESNASVIFDYSDDGRLIKTSFLTGTGRQV 1977  
Db 1889 DYLLSVTPSPSVRHALQMLSVGYRNIYTPPDSGAFTQDVTDRCLLQTLTLPYGTGRV 1948  
Qy 1978 FYKYKLSKLSIEVYDSTAVTFGYDETGVKLMVNLQSGFECTIRYRKIGPLVQKLYR 2037  
Db 1949 LYKYSQKRLSEILDYDTQVTFTEESSGVKTIHLMHDGFICTIRYKOTGPLIGRQIFR 2008  
Qy 2038 FSEGMVNAFYDTHONSFRIASIKPVISETPLVDLYRYDEISGKVEHFKGFIYVD 2097  
Db 2009 FSEGLVNAFYDYSI--NNFRVTSQAMINETPLDIDLYRYVDVSGRTEQFGKFSVINYD 2066  
Qy 2098 INQIITAVMTLSKHFDTGRIKEVOYEMFRSLMYMTVQVDSMGRVTKRELKLPYANT 2157  
Db 2067 LNOVITTVMKTKIFSAANGOVIEVOYEILKSIAYWMTIQYDNMGRVVICDIRGVGANI 2126  
Qy 2158 TKYTYDYGDOGLOQSVAVANDRPTWRYSYDLNGLNLLNPGNSVRLMPLRYDLDRITRLG 2217  
Db 2127 TRYFYEDRGOLOQVSYNDKTQWRYSYDLNGLNLLSHGNSARLTPLRYDLDRITRLG 2186

```
QY 2218 DVQYKIDDDGYLCQSGSDIFEYNKSGLLTRAYNKASGWSVQVRYDGVGRASRYKTNLGH 2277
Db 2187 EIQQKDEDFLQRQNEIFEYNSGLNKKAYNKVSGWTVOYCYDGLRRVASKSLGQH 2246
QY 2278 LOYFVSLDANPRITHVYHNSSEITSLAYDLOGHLFAMESSGGEYVVASDNTGTPLAY 2337
Db 2247 LOFFYADLSNPRIVTHLYHNSSEITSLAYDLOGHLIAMELSSGGEYVYACDNTGTPLAV 2306
QY 2338 FSTNGLMIKOLOYATYGEIYDSDNPDQFMVIGFPHGGLYDPLTKLVHFTQRDQYDVLAGRWT 2397
Db 2307 FSSRGQVKEILYTPYGEIYQDTPDQFVVIGFPHGGLYDLSLTKLVHLOGRDYDVIAGRWT 2366
QY 2398 SPDYTMKWKVGEKPEAPFNLYMFKSNPLSSELDLKNYVTDVKSWLVMFQFQSLNIIPGRP 2457
Db 2367 TPNHHIWKHLNAVQPFNLYSPENNYIQDIAKVTITDIGSWLELFGQLHNVLPFGP 2426
QY 2458 RAKMYVPVPPYEL---SSQASENQLITGVQOOTERHNOAFMALEGQVITK----KL 2508
Db 2427 KPEIEALETTYELLQOTKQWDPQKTLIGIQELQKQLRNFISLDQLPMTPRYSDGKC 2486
QY 2509 HASIREKAGHWFATTPPIICKGIMFAIKGRVTVTGSSIASDSRKAVASVLNNAYYLDKM 2568
Db 2487 YEGVKQPR--FAAIPSVFGKIGKFAIKDGIYVADIIGVANEDSRRIAAILNNAHYLENL 2543
QY 2569 HYSIEKQTHYFKVIGSADGLVTLGTTIGRKVLESGVWTVVSQPTLLVNGRTRRTNTE 2628
Db 2544 HFTIEGRDTHYFKIGLSLEEDLSLIGTGGRILENGVWTVVSQMTSVINGRTRRPADIQ 2603
QY 2629 FOYSTLLSIRYGLTPDTLDEKARVLDQARORALGTANAKQOQKARDREGSRWLWTEGE 2688
Db 2604 LHGALCFNVRG---TTVEEKNHVLEVARQAVAAQATKQORRLQOEBEGIRAWTDGE 2660
QY 2689 KOOLLSTGRVQYEGYVILVPEYQPELADSSNIQFLRQNMGRK 2733
Db 2661 KOOLLNTGRVQYDGYFVLSVEQYLELSDSANNHPRMQSEIGRR 2705

RESULT 12
Q9WTS4
ID Q9WTS4 PRELIMINARY; PRT: 2731 AA.
AC Q9WTS4;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE TEN-M1.
GN OD21 OR TEN-M1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
STRAIN=BALB/C;
RA Oohashi T., Zhou X., Feng K., Richter B., Moergelin M., Perez M.T.,
RA Su W., Chiquet-Ehrismann R., Rauch U., Faessler R.;
RT "Mouse Ten-m/Odz is a new family of dimeric type II transmembrane
RT proteins expressed in many tissues.";
RL J. Cell Biol. 0:0-0(1999).
DR EMBL: AB025410; BAA77396.1; -.
DR HSSP: P00750; 1TPG.
DR MGD: MGI:1345185; Odz1.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR001258; NHL.
DR InterPro: IPR000209; Peptidase_s8.
DR Pfam: PF00008; EGF; 5.
DR Pfam: PF01436; NHL; 3.
DR PRINTS: PRO0011; EGF_LAMININ.
DR SMART: SM00181; EGF; 7.
DR PROSITE: PS00022; EGF_1; UNKNOWN_8.
DR PROSITE: PS01186; EGF_2; 7.
DR PROSITE: PS00136; SUBTILASE ASP; UNKNOWN_1.
KW EGF-like domain; Glycoprotein.
```

```
SQ SEQUENCE 2731 AA: 305792 MW: 9129FA4CFE4A7770 CRC64;

Query Match 58.2%; Score 8484.5; DB 11; Length 2731;
Best Local Similarity 55.9%; Pred No. 0;
Matches 1561; Conservative 476; Mismatches 618; Indels 139; Gaps 22;

QY 8 HRSLTRGRCGKCRITYTSSSLSDSCRPVTKQKYSYSSSETLKAYDHDHSDRMHYGNRVDTLIHR 67
Db 9 YQPLSKVKHEMDLAYTSSSEDESDGRKPRQ-SFNSRETLHEYNOELRRV----- 57
QY 68 ESEDFPRQGTNTLAELGICEPSPHR-SGYCDM-CILHQGYSLSTGSDADSDTEGMSR 125
Db 58 NSQSRKRKDKVEKSTQIEFECPTPTLCGYHTDMHSVSRHGYLEMGSDVDTEGAASP 117
QY 126 EHAIRLWGRGIRSRSSGLSSRENSALTITDSDNEKNKSDDENGRPIPTTSSPSLLPSAQL 185
Db 118 DIALRWIIRGMKSEHSSCLSSRANSALSITDTHERRKSDGNGFKF----- 163
QY 186 PSHNPPVSCOMPLDLSNTSHQIMDTPDEEFSNYSLLLRACSGPQQAASSGPPNHHSQ 245
Db 164 -----SPVCCDM-----EAPADSAQDMQSSPHNQFTF 190
QY 246 STLRPLPLPPPHNHTLSHH-HSSANSLNRSLNRRSQIHAPAPAPNDLATTPEVOLQDS 304
Db 191 RPLPPPPPPHACTCARKPPTVDLSLRKSMWT-RSQSPAAPAPP--TSTQDSVHLHNS 247
QY 305 WYLSNVNPLETRHFLFKTSGSTPLFSSSPGYPTLSGTVTPPPRLPLNPTFSRKAFL 364
Db 248 WYLSNINPLETRHFLFKHSGSSAIFSAASONYPLTSNTVYSPPPRPLRSFTRPAFTF 307
QY 365 KPSKYCSWKAALSIAAALLAILLAYFI-----VPWSL----- 400
Db 308 NKPYRCNNKCTALSATITVTLLALLAYVIAVHLFGLTWLQPVGOIYANGISNGPST 367
QY 401 -----KNSSIDSGEAEVGRVTOEVPVGVWRSGIHSISQO 436
Db 368 ESMDTTYSPIGRVSDKSEKVFQKRAIDTGEVDIGAQMOTIPPLGLFWRRQITTHFI 427
QY 437 FLKENISLGKDALFGVYIRRGPPSHQAQYDFMERLDKE-----KWSVVESSPRERS 488
Db 428 YLKENISLAKDSLGIYGRNPTPTQTFDFVKLMDGKQLVKQDSKSSDIQHSR--RNL 485
QY 489 IQTLVONEAVFQYLDVGLWHLAFYNDGDKKEMVSENTVVLDSVQDCPRCHNGSGCVS 548
Db 486 ILTSQ-ETGFIEYMDQGPWYLAFYNDGKMEQVFLTAIEMDDCSTNCNGECISG 544
QY 549 VCHCFPGFLGADCAKAAACPVLCSGNGOYSGTKOCYSGNKGAECVPMNQCIDPSCGGHG 608
Db 545 HCHCFPGFLGPCARDSCPVLCGGNGEYKGHCVCRNGKNGKGPEDVPEQCIDPTCFHG 604
QY 609 SCIDGNVCVSAGYKGEHCEBEVDCLDPTCSSHGVYVNGECLCSFGWGLNCELARVQCPQ 668
Db 605 TCIMGVICVCPYKGEICEEDCLDPMCSHGIYKGECHCSTGMGVNCEPTLPICQEQ 664
QY 669 CSGHGTYPDTLGLSCDPPNMGPDSCSVFVSCVDCGTHGVICGACRCBEGWGAACDQRV 728
Db 665 CSGHGTFLDITGVCSDDPKWTGSDCTELCTWECGSHGVCSRGICQCEBEGWVGTPEERS 724
QY 729 CHPRCIEHGTCKDGKCECEGNGEHCTIGROTACTGTGCPDCLNGNCRCTLGQNSWOC 788
Db 725 CHSHCAEHGQCKDGKCECEGNGEHDCTIAHYLDVAVR-DGCPGLCRFGNCRCTLDQNGWHC 783
QY 789 VCQTGWRGPGCNVAMETSCADNKNEDGLVDCLDPCCLQACQNSLRCGRSRDPLDII 848
Db 784 VCOVGWGTGCVNVMELCGDNLNDNDGGLTDCVDPDCCQSNVCYSPLCQSGSPDLDI 843
QY 849 QOQOTDWPVKV--FYDRIKLAGKDSHIIICGENFENFNSLSVSLIRGQVVTDTGTPLVG 906
Db 844 QOQPLFSQHTSRLEFYDRIKFLIGDKDTHVVPQDISFSDRSRACVIRGQVAVDGTPLVG 903
QY 907 NVSFVKYPKYGTITRODCTFDLIANGASLTLLHFERAPFMSQERTVWLPWNSFYAMDLT 966
Db 907 NVSFVKYPKYGTITRODCTFDLIANGASLTLLHFERAPFMSQERTVWLPWNSFYAMDLT 966
```

Db 904 NVSFLHSDYGTISRQDGSFDLVAIGGISVLFIDRSPFLSEKRTLWLPWNQVIVVEKV 963  
Qy 967 VMKTEENSIPSCDLSGFRVPPPIIISPLSTSFSAAPQONPIVPTOVLEHEIEELPGSNV 1026  
Db 964 IMQIVADAPSCDISNFISPNPIVPLSPFTSGGSCPRGTIVFELQVQQEIPSPSFV 1023  
Qy 1027 KLRYLSSRTAGYKSLTKITMTOSTVPLNLIRLHVLMVAVEGHFLQKSFQASPNLSTFTWD 1086  
Db 1024 RLSYLSRSRTPGYKTLRLTLRLLTHSTIPVGMKIVHLTVSVEGRLTQKWFPAALNLVYTFAWN 1083  
Qy 1087 KTDAYGQRYVGLSNAVSVSGEYETCPSLILWEKRTALLOFELDPSPNLGWSLDDKHHIL 1146  
Db 1084 KTDYIGQKRWGLAEALVSGVEYEMCEPFIWEQRTVYVLOGFEMDASNLGWSLKNHHIF 1143  
Qy 1147 NVKSGILHKGTEGNOFLTQOPAITTISMGNGRRRSISCPSCNGLAEGKNLLAPVALAVGI 1206  
Db 1144 NPGSGIIHKNGENWFISQPPVATIMNGHQHSVACTNCGPAHNNKLPAPVALASGP 1203  
Qy 1207 DGSLYVGDNFYIRIFPSPNVTISLELRNKEFKHNNPAHKYILAVDPVSGSLXVSOTNS 1266  
Db 1204 DGSVYVGDNFVRIFFPSGNSVSIELLRNDRTHSTSPAHKYILAMDPMSESLYLSOTNT 1263  
Qy 1267 RRIYRVKSLCTKLAGNSEVYVAGTGEQCLPFDEARCGDGKADATLMSRGLTAVDKNG 1326  
Db 1264 RKVYKLSLVETKLSKNFEVYVAGTGDQCLPFQDGHGDKGKASEASLNSRPGITVDRHG 1323  
Qy 1327 LMVYFVDMIRKVPDQNGIISTLLGSLNDLTAVRPLSCDSSMDVAOVRLEWPTDLAVNPMDN 1386  
Db 1324 FIYFVDMIRIDENAVITVIGSNGLTSTQPLSCDSGMDITQVRLEWPTDLAVNPMDN 1383  
Qy 1387 SLYVLENNVILRITENHOVSIIAGRPMHCQVPGIDYSL- SKLATAHSALESASAIASHTG 1445  
Db 1384 SLYVLDNNVILQISENNRVRRIIAGRPICQVPGIDHFLVSKVAITHSTLESARAISSVSHG 1443  
Qy 1446 VLYITETDEKKINLRQVTTNGEICLLAGASDCDCKNDVNCNCSGDDATATDAILNSP 1505  
Db 1444 LLFIAETDERVKNRIQVTTNGEISIIAGAPTDCDCKIDPNCDFSGDGGYAKADAKMKAP 1503  
Qy 1506 SSLAVAPDGTIADLGNIRIRAVSKNPVLNFAFNQYEAASPGEOELVFNADGTHQVTV 1565  
Db 1504 SSLAVSPDGLVADLGNVIRITISKNOHLNDMNLIEIASPADELYQFTVNGTHLTM 1563  
Qy 1566 SLVGEYLYNTYSTDNDVTDELIDNNGSLKIRDDSCGPHRLMDNQITLTVGTTNGG 1625  
Db 1564 NLITRDYVYNYETNAEGDLGAIITSSNGSNVHIRDAGGMLVWVPGGVYWLATSSNGV 1623  
Qy 1626 LKVYSTONLEGLMTYDGTGTLATKSDGTFTFYDYDHEGRLTNVTRPTGVVTSLSHRE 1685  
Db 1624 LKRVSAQGYNLALMTPYDGTGTLATKSNENGWTTVYETDPEGHLTNATFTTGEVSSFSHD 1683  
Qy 1686 MEKSTIDIENSRDDDDVTYITNLSSVEASTYVQDOVRNSYQLCNCNGTLRVMYANGMI 1745  
Db 1684 LEKLTKVADLTSNR-ENVLMSTNLATSTIVILKQENTQSTYRVSPDGSRLRVTFSAGMEI 1742  
Qy 1746 SFHSEPHVLAGTPTTIGRCNLSLPMENGLNSIENRLEKQIKGVITFGRKLRVHGNL 1805  
Db 1743 NLSSEPHLAGAVNPFTLGKCNISLPGERNANLIEWRQKEQNGKVSAPAFERRLAHNRL 1802  
Qy 1806 LSIDYDNRIRTEKIIDHDKFTLRIYDQVGRPFLLWLPSSGLAAVNVSYFENGLRAGLQR 1865  
Db 1803 LSIDFDMHTRTKIYDHRKFTLRLYDQTRPILWSPVRYNEVNTIYSPSGLVTFYFQR 1862  
Qy 1866 GAMSERTIDKOGRIVSRMFADGKWSYSYLDKSMVLLQSORQYIFEYDSSDRLLAVTM 1925  
Db 1863 GTWNEKMEYDQSGKIISRTWADGKIWSYILEKSVMLLLHSQRRYIFEYDQSDCLLSVTM 1922  
Qy 1926 PSVAHRSMTHTSGYIRINIVNPPESNASVIFDYSDDGRILKTSPLTGROVYFYKGLS 1985  
Db 1923 PSWVRHSLQTLMSVGYRYNIYTPDSTSFIDQYSRDRGLLQTLHLGTGRRVLYKYTKQA 1982  
Qy 1986 KLSIETVYSTAVTFCYDETTGVKLMVNLOSQSGFCTIYRKIGPLVDKQIYRFSEGMVN 2045  
Db 1983 RLSEILYDTTQVTLTYEESSGVIKTIHLMHIDGFICTIYRQTGPLIGRIYRFSEGLVN 2042

Qy 2046 ARFDYTYHDNSFRIASIKPVISETPLPVDLYRYDEISGKVEHFGVGYIYDINOIITTA 2105  
Db 2043 ARFDYSY--NNFRTVSMQAVINETPLPIDLYRYDVDSORTQFQKFSVINDLQNOVITT 2100  
Qy 2106 VMTLSKHFDTHGRKEVQYEMFRSLMYMTVQYDSMGVRVIRKELKGLPYANTKYTYDYD 2165  
Db 2101 VMKTKIFNANGQVTEVQYELKAIAVMTIYQDNMGRVICTDIRVGDVANTIRYFYEYD 2160  
Qy 2166 GDGQLOSVAVNDRTWRYSYDLNGNLHLLNPGNSVRLMPLRYDLRDRITRIGLDVQYKIDD 2225  
Db 2161 ADGQLOTYSVNDKQWRYSYDLNGNLHLLNPGNSARLTPLRYDLDRTIRLGEIQYKNDE 2220  
Qy 2226 DGYLCQSGSDTIFEYNSKGLLTRAYNKASGWSQVRYDCVGRRASVTKTLNGLHLLQYFYSD 2285  
Db 2221 DGFRLRQGRNDLFEYNSGGLLQKAYNKVSGWTQVYYDGLGRVASKSLSGOHLQFFYADL 2280  
Qy 2286 HNPTRITHVYHNSSETSLYDLOGHLFAMESSGEYVYVASDNTGTPLAVFSINGLMI 2345  
Db 2281 ANPIRVTHLYNHTSAEITSLYDLOGHLIAMELSGSEYVYVACDNMGTPLAVFSRQGV 2340  
Qy 2346 KQLOYTAYGEIYDSDNPDQMVIGFPHGLYDPLTKLVHFTORDYDVLAGRWTSPDYTMWK 2405  
Db 2341 KEILYTPGDYHDYDTPDEFVIGFPHGLYDPLTKLVHGLQRDYDVVAGRWTTPNHHLWK 2400  
Qy 2406 NVGKEPAPNLYMFKSNPLSSELDLKNYVTDVKSWSLVYMFQFQSLNSIIPGFPRAKMYFVP 2465  
Db 2401 QNLNLLPKPFNLYSPENNYPVGKIQDVAKYTTDGTWLELFGQLHNLVLPFPKPELENWE 2460  
Qy 2466 PPYEL-----SSQASENGQITGVQOOTTERRINQAPMALEGQVITKKLHASREKAGHW-- 2519  
Db 2461 LTYELQLQOTKTDQWDPGKMILGQCELOKQLRNFISLDQLPMPQYNEGRGLEGGKQPR 2520  
Qy 2520 FATITPIIGKIGMFAIKRGVTTGVSSTASEDSRKSASVLANNAYVLDKMHYSIEGKDPHY 2579  
Db 2521 FAAPVSPGKIGKFAIKREGIVTADIGVANEDESRKLAAILANNAHLEHLEHTIEGRDTHY 2580  
Qy 2580 FVKIGSADGLVLTGTTIGRKVLESQVNVTVSQPTLLVNGRTRFTNIEFOYSTULLSIR 2639  
Db 2581 FIKLSLEEDLVLTGTTGRRILENGVNVTVSQMVSVLNGRTRFRADLOLHGALCFENIR 2640  
Qy 2640 YGLPDPDLDEKARVLDQARALGTAWAKEQKARDREGSRLWTEGEKQOLLSTGRVQ 2699  
Db 2641 YG---TTVEEKNHYLEMARQAVQAWTOEORLQEGEGTRVWTEGEKQOLLSTGRVQ 2697  
Qy 2700 GYEGYVLPVQYPELADSSNIQFLRONEMGR 2733  
Db 2698 GYDGYFVLSVEQYLELSDSANNIHPMROSEIGRR 2731

RESULT 13

Q9UKZ4

ID Q9UKZ4

AC Q9UKZ4

DT 01-MAY-2000

DT 01-MAY-2000

DT 01-DEC-2001

DE TENASCIN-M1

GN TNM1

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20025749; PubMed=10556288;

RA Brandau O., Schuster V., Weiss M., Hellebrand H., Fink F.M.,

RA Kreczy A., Friedrich W., Strahm B., Niemeyer C., Belohradsky B.H.,

RA Meindl A.;

RT "Epstein-Barr virus-negative boys with non-Hodgkin lymphoma are

RT mutated in the SH2D1A gene, as are patients with X-linked

RT lymphoproliferative disease (XLP).";

RL Hum. Mol. Genet. 8:2407-2413(1999).

DR	EMBL: AF100772; AAF04723.1; --	
DR	HSP; P00750; ITPG.	
DR	InterPro: IPR000561; EGF-like.	
DR	InterPro: IPR002049; Laminin_EGF.	
DR	InterPro: IPR001258; NHL.	
DR	InterPro: IPR000209; Peptidase_S8.	
DR	Pfam: PF00008; EGF; 5.	
DR	Pfam: PF01436; NHL; 3.	
DR	PRINTS: PR00011; EGF_LAMININ.	
DR	SMART: SM00181; EGF; 6.	
DR	SMART: SM00001; EGF-like; 1.	
DR	PROSITE: PS00022; EGF_1; UNKNOWN_8.	
DR	PROSITE: PS01186; EGF_2; 7.	
DR	PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.	
KW	EGF-like domain; Glycoprotein.	
SQ	SEQUENCE 2725 AA; 304892 MW; 6F383E64C1B2E1B1 CRC64;	
Query Match 57.7%; Score 8406.5; DB 4; Length 2725;		
Best Local Similarity 56.2%; Pred. No. 0;		
Matches 1562; Conservative 454; Mismatches 619; Indels 145; Gaps 23;		
Db	22 YTSSSLSDECRPTOKSYSSSETLKAYDHSRMHYGNRVTLIHRESDEEPROGTFNL 81	
Db	23 YTSSSDESEDRKPRO-STNSRETLHEYNOELRMNT-----NSOSRRKKEVEKST 71	
Qy	82 AELGICEPSPHR--SGYCSDM-GILHQYSLSTGSDADSTDEGMSPEHAIRLMGRGKS 138	
Db	72 QEMEFCEFS-HTLCSGYQTDMHSVSRHGYOLEMGSDVDTETEGAAASPDHALRMWIRGMS 130	
Qy	139 RRSGLSSRENSALITSDSDENKSDDENGRIPTTSSPSLPSAQLPSHNPVPSVCOM 198	
Db	131 EHSCLSSRANSALSITDTHERRKSDGNGFKF-----SPVCCOM 170	
Qy	199 PLLDSNTSHQIMDTNPDEEFPNSYLLRACSGPQQAASSGPPNHHQSQSLRPLPPPHNH 258	
Db	171 -----EAQAGSTQDVQSSPHNQFTFRPLPPPPPHAC 203	
Qy	259 TLSSH-HSSANSLNRLNRSQIHAPAPNDLATTPEVOLQDSWVLSNVPLETRH 317	
Db	204 TCARKPPAAPASLQRRSMTT-RSQPSAAPAPP--TSTQDSVHLHNSWVLSNVPLETRH 260	
Qy	318 FLFKTSSGTPLEFSSSPGYPITSGVYTPPPRLLPRNTFSKAFKLKPSKYSCKKCA 377	
Db	261 SLFKHSGSSAIFSAASQNYPLTSNTVYSPPPRPLPRSTFSRPAFTFNKPYRCNNKCTA 320	
Qy	378 LSAIAALLAILLAYFI-----VPMSL----- 400	
Db	321 LSATATVTLALLAIVIAVHLFGLTWQLOPVEGELYANGVSKNGRGTESMDTTYSPIG 380	
Db	401 -----KNSSIDSGAEVGRVTTQEPVPGVFWRSQIHISQPFLLKFNISLGKDA 448	
Db	381 KVSDDKSEKKVFOGRAIDTGEVDIGAQVMQITPPGLFWRFQITIHPIYLKFNISLAKDS 440	
Qy	449 LFGVYTRGLPSHAQYDPMERLDKE-----KNSWSPRERRSIQTLVONEAVFOY 502	
Db	441 LGLGIYRRNIPPTHQTFDEFLKMDGKQLVKQDSKGSDDTQHSRNLILTSLQ-ETGFIEY 499	
Qy	503 LDVGLWHLAFYNDGDKEMVSENTVVLDSVDQPCPRNCHGNGECVGVCHCFPGFLGADCA 562	
Db	500 MQQGPWYLAFYNDGKMEQVFLVTLTAIETMDDCSTNCNGECISGCHCFPGFLGADCA 559	
Qy	563 KAACPVLCSNGQYSGKTCCQYSGWKGAECVPMNQIDPCSGGHGSCIDGNCVCSAGYK 622	
Db	560 RDSCPVLCGGNGEYEGKHCVRHNGKGPEDVPEEQIDPTCFGHCITMGVCICVPYK 619	
Qy	623 GEHCEVDCLDPTCSSHGVYNGECLCSPGWGLNCELARVQCPDQCSGHGYLPTDGLC 682	
Db	620 GEICEEDCLDPMCSNHGTCVGECHCSTGWGGVNCETPLPYVCQEQCSGHGTFLLDAGVC 679	
Qy	683 SCDPNMGMGDSVEVCSVDGTHGVICGACRCEEGWGAACDQQRVCHPRCTIEHGTCKDG 742	
Db	680 SCDPKWTGSDCTELCTMECGSHGVCSRGICQCEEGWVGPTCEERSCHSCTEHGOCKDG 739	

Qy	743 KCECREGMNGEHCTIGROTAGTETDGCPLCNGNGRCTLGQNSWCVCYCTGHRGPCNVA 802	
Db	740 KCECSPWEGEDHCTIAHYLDAVR-DGCPGLCFGNGRCTLQNGWHCVCGVSGTGCNVV 798	
Qy	803 METSCADNKDNEGDGLVDCLDPDCCLOSAQNSLLCRGSRDPLDIIQOQOTDHPAVKS-- 860	
Db	799 MEMLCGDNLDNDGDLTDCVDPCCOOSNCYISPLCOGSPDPLDLIIQOQSOTLFSQHTSRL 858	
Qy	861 FYDRIKLAGKDSITHIPGENFENSSLSVLIRGQVYTTGCTPLVGVNVSVFKYKGYTI 920	
Db	859 FYDRIKFLICKDSTHVIPEVFSFDSRRACVIRGQVVAIDGCTPLVGVNVSVFLHSDYGFTI 918	
Qy	921 TRQDCTFDLIANGGASLTLLHFERAPPMQSBERTVWLPWNSFYAMDTLVLMKTEENSIPSCDL 980	
Db	919 SRQDGSFDLVAIGGISVILIFDRSPFLPEKRTLWLPLWPNQFIVVEKVTMQRVVSDPPSCDI 978	
Qy	981 SGFVRPDPPIIISPLSTFFSAAPGNPIVPEQVLRHEEELPGSNVVKLYLSRRTAGYKS 1040	
Db	979 SNFISNPILVLSPLTSFGSCSPERGTIYVPELQVQVEEIPIPSSFVRLSYLSRTPGYKT 1038	
Qy	1041 LKIKMTQSTVPLNLIRVHLMAVEGHFKQKFOASPNLASTFIMDKTDAYGORVYGLSD 1100	
Db	1039 LRLILLTHSTIPVGMIKVHLTVAVEGRLTQKWFPAINALVYTFANWKTIDYQKQWGLAE 1098	
Qy	1101 AVYGVGFYETCPSLILWEKRTALQGFELDPNSLNGWGLDKHHILNVKSGILHKGTGN 1160	
Db	1099 ALVSVGYEYETCPDFTLWEQRTVVLQGFEMDASNLGDKHHLNPOSGIHKNGEN 1158	
Qy	1161 QFLTQOPALITSIMGNRRRSISPCSCNGLAEGNKLLAVALAVGIDGSLYVCDNRVIR 1220	
Db	1159 MFIQQPPVISTIMGNHQRSVACTNCNGPAHNKLFAPVALASGPGSVYVGDGFVRR 1218	
Qy	1221 IFFSRNVTILELRNKEFKHSNNPAKYYLAVDPVSGSLYSDVTSNRRIRYRVKSLGTXD 1280	
Db	1219 IFFSGNSVILEL-----STSPAHHYLLAMPVSESLYSDTNTRYVYKLSLVETKD 1271	
Qy	1281 LAGNEVAVAGTGEQCLPPDEARCGGGKAIADATLMSPRGIAVDKNGLMTFVDMTRKYD 1340	
Db	1272 LSKNFEVAVAGTQDCLPPDQSCGCGGRASEASLNSPRGITVDRHGFIYFVDGTMIRKID 1331	
Qy	1341 QNGIISTLLGNDLTAVRPLSCDSSMDVAQVLEWPTDLAVNPMDSLVLENNVLTIT 1400	
Db	1332 ENAVITVIGSNGLTSTQPLSCDSDGMDITQVRLWEPTDLAVNPMDSLVLENNVLTIT 1391	
Qy	1401 ENHQVSIAGRMHCOVPGIDISL-SKLAIHSALESASAIASHTGLVITITDEKKINR 1459	
Db	1392 ENRVRIIAGRIHCOVPGIDHFLYSKVAIHSTLESARAISSHGLLIAETDERKVN 1451	
Qy	1460 LRQVTTNGEICLLAGAADCDCKNDVNCYSGDDAYATDAI LNSPSSLA VAPDGTIYTA 1519	
Db	1452 IQQVTTNGEYIYIAGAPTDCDKIDPNCDFSGDGGYAKDAKMAKPSLAVSPDGTLYA 1511	
Qy	1520 DLGNIRIRAVSNKPVNAFNOYEAASPGEOELVFNADGIIHQYTVSLVTGVELYNFTYS 1579	
Db	1512 DLGNVRIITSRQAHLNDMNIYEIASPADQELQVTVNGTHTLNLITRIDYVYVNTYN 1571	
Qy	1580 TONDVTELDNNGNSLKIIRDSSGMPRHILMPDNQIITITVGTNGGLKVVSTQNLEGLM 1639	
Db	1572 SEGDLGAISSNGNSVHIRRAGGMLWLVPVGGQVYVLTISSNGVLKRVSAQGNPALM 1631	
Qy	1640 TYDQNTGLLATSDETGWTTFYDYDHEGRLTNVTRPTGVVTSLHREMEKSIIDENSNR 1699	
Db	1632 TYPGNTGLLATKSNENGWTVYEDPEGHLTNATFPTGEVSFHSFSLDEKLTVELDTSNR 1691	
Qy	1700 DDVTVITNLSSVEASYTVVQDVRSYOLCNGTLNRVYANGMGISPSHPIVLAGTTT 1759	
Db	1692 -ENVLMSTLNTATSTIYILKQENTGSTYRNPDGSLURVTFASGMEIGLSEPHILAGAVN 1750	
Qy	1760 PTIGRCNISLPMENGLNSTEWLRKEIQKGYTIFGRKLVRHGRNLLSIDYDRNTRTERKI 1819	
Db	1751 PTLGKCNISLPGEHANLIEWQRKEQNGYSAFERRLRHNRNLLSIDFDHITRTGKI 1810	

Qy 1820 YDHRKFTLRIYDQVGRPFLWSPSSGLAAVNVSYFFNGRLAGLQRGAMSBRTDIDKQGR 1879  
Db 1811 YDHRKFTLRIYDQVGRPFLWSPVSYFFNGRLAGLQRGAMSBRTDIDKQGR 1879  
Qy 1880 IVSRMFADGKWSYLDKSWLLQSQRYIFEDSSDRLLAVLMPFSAVHSMSTHTSI 1939  
Db 1871 IISRTWADGKWSYLDKSWLLQSQRYIFEDSSDRLLAVLMPFSAVHSMSTHTSI 1939  
Qy 1940 GYININYPSPASVDFDSDGRILKTSFGLGRGVFKYKGLSKLSEIVDYDSTAVTF 1999  
Db 1931 GYININYPSPASVDFDSDGRILKTSFGLGRGVFKYKGLSKLSEIVDYDSTAVTF 1999  
Qy 2000 GYDFTGVLKMNVLQSGFGCTIRYKIGPLVDKQIYFSEEGMVNAREFYTHDNFSRI 2059  
Db 1991 TYEESGVIKTHLMDHGFCTIRYKIGPLVDKQIYFSEEGMVNAREFYTHDNFSRI 2059  
Qy 2060 ASIKPVLSEPLVDLYRYDEISKVEHFGKGVYIYDINOIITAVMTLSKHEDTHGRI 2119  
Db 2049 TSMQAVINETPLPDLRYVDVDSRTQFQKFSVINDLQVITTTVNMKHTKIFSANGQV 2108  
Qy 2120 KEVOYEMFRSLMYMTVQYDSMGVIRKELKPGYANTTKYTYDYDGDGQLQSVAVNDP 2179  
Db 2109 IEVOYEILKATAYMTIQYDNGVRHNCIRGVYDANITRYFEYDADGQLQTVSYNDKT 2168  
Qy 2180 TWYSDYDLNGLHNLNPGNSVRLMPLRYDLRDRITRLGDVQYKIDDDGYLCQSGDIFEY 2239  
Db 2169 QWYSDYDLNGLHNLNPGNSVRLMPLRYDLRDRITRLGDVQYKIDDDGYLCQSGDIFEY 2239  
Qy 2240 NSKGLLFRAYKASGWSVOYRYDGVGRASYKTNLGHHLQYFYSDLHNPTRITHVYHNSN 2299  
Db 2229 NSNGLLQAYKASGWSVOYRYDGVGRASYKTNLGHHLQYFYSDLHNPTRITHVYHNSN 2299  
Qy 2300 SEITSLYDLOGLHFLAMESSEGEYVYASDNTGTPLAVFSINGLMIKOLOYATYGETYYD 2359  
Db 2289 SEITSLYDLOGLHFLAMESSEGEYVYACDNTGTPLAVFSINGLMIKOLOYATYGETYYD 2359  
Qy 2360 SNPDFQVIGPHGLYDPLTKLVHFTQRYDVLGRMTSPDYTWKNNVKGEPAPFNLYMF 2419  
Db 2349 TYPDFQVIGPHGLYDPLTKLVHFTQRYDVLGRMTSPDYTWKNNVKGEPAPFNLYMF 2419  
Qy 2420 KSNPLSSELDKLVYDVKSWLVMFGQLSNIIPGPRAKMYFVPPPYEL---SESQA 2475  
Db 2409 ENNTPVGIQDVAKYTTDIRSWLELFGQLHNVLPFGPKPELENLELTLYELLRLQTKQE 2468  
Qy 2476 SENGQITGVQOITERRHQAFLMALEGOVYTKKLHSAIREKAGHW--FATTTPIIGKIMF 2533  
Db 2469 WDPCKTILGIQELQKOLRNFIISLDQLPMTPRYNDGRCLGGKQPRFAAVPSVFGKIKF 2528  
Qy 2534 AIKGRVTTGVSSTASEDSRVSASVNLNAYYLDKMHYSIEGKDRTHYFVKIGSADGLVTL 2593  
Db 2529 AIKGRVTTGVSSTASEDSRVSASVNLNAYYLDKMHYSIEGKDRTHYFVKIGSADGLVTL 2593  
Qy 2594 GTTIGRKVLSEGVNVTYVQPTLLVNGRTRFRFTNIEFOYSTLLLSIRYGLTPTDIDEKAR 2653  
Db 2589 GNTGGRILEGVNVTYVQPTLLVNGRTRFRFTNIEFOYSTLLLSIRYGLTPTDIDEKAR 2653  
Qy 2654 VLDQARQALGATAWAKQEQKARDREGSRRLTEGEKQQLSTGRVQGVYEGYVLPFVEQYP 2713  
Db 2646 VLEIARQAVAAQWTKQRRRLQEGEGRIRATEGEKQQLSTGRVQGVYEGYVLPFVEQYP 2705  
Qy 2714 ELADSSNIQFLRONEMGKR 2733  
Db 2706 ELSDSANNIHPMRQSEIGRR 2725

RESULT 14  
Q9P273 ID Q9P273 PRELIMINARY; PRT: 1769 AA.  
AC Q9P273; 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE KIAA1455 PROTEIN (FRAGMENT).

GN KIAA1455.  
OC Homo sapiens (Human).  
OC Eukaryota: Chordata: Craniata: Vertebrata: Euteleostomi;  
OC Mammalia: Eutheria: Primates: Catarrhini; Hominiidae; Homo.  
OX NCBI\_taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20277482; PubMed=10819331;  
RA Nagase T., Kikuno R., Ishikawa K., Hirotsawa M., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human  
RT genes.XVII.The complete sequences of 100 new cDNA clones from brain  
RT which code for large proteins in vitro.";  
RL DNA Res. 7:143-150(2000).  
DR EMBL; AB040888; BAA95979.1; .  
FT NON TER 1  
FT NON TER 1769 1769  
SQ SEQUENCE 1769 AA; 199209 MW; 2E1524DC90F29125 CRC64;  
  
Query Match 46.8%; Score 6822; DB 4; Length 1769;  
Best Local Similarity 70.6%; Pred. No. 0;  
Matches 1255; Conservative 255; Mismatches 248; Indels 20; Gaps 5;  
  
Qy 833 QNSLLRCGRSDPLDIIOQG--QTDWPAVKSYDRIKLAGKDSHTIIPGENPFFSSLSVL 890  
Db 1 QNQPCYCRGLPDPDIIISQSQSPQQAQKSYDRISFLIGSDSTHVIPGESPFNKSLASV 60  
Qy 891 IRGOVVTDTGTPLVGNVSVFKYKGYTITRODGTDLIANGCASLTLLHFERAPFMSOE 950  
Db 61 IRGOVLTADGTPGLGVNVSFHYPEYGTITRQDMFDLVANGASLTLLVFERPFLFYQ 120  
Qy 951 RTVWLPNWSFYAMDTLVMKTEENSIPSCDLSGFVRPDPPIISPLSTFFSAAPQGNPIV 1010  
Db 121 HTVWLPNWSFYAMDTLVMKTEENSIPSCDLSGFVRPDPPIISPLSTFFSAAPQGNPIV 180  
Qy 1011 ETQVLHBEIEPLGPNVRLYSSRTAGYKSLKLTWTQSTVPLNLRVHLVAVEGHFLQ 1070  
Db 181 ETQVLHBEIEPLGPNVRLYSSRTAGYKSLKLTWTQSTVPLNLRVHLVAVEGHFLQ 240  
Qy 1071 KSFOASPLASTFTWKTDAYQGVYGLSDAVSVSGVEYETCPSLILWEKRTALQSGEL 1130  
Db 241 KWFASPENLAVTFWKTDAYQGVYGLSDAVSVSGVEYETCPSLILWEKRTALQSGEL 300  
Qy 1131 DPSNLGGWSLKHILNVKSGILHKGTGENOFLQOQPAITISIMNGRRRISCPSCNGL 1190  
Db 301 DASNMGGWTLKHVLDVQNGILYKNGENOFISQOQPPVSSIMNGRRRISCPSCNGL 360  
Qy 1191 AEGNKLAPVALAVIDGSLVVGDFNYTRIRFPSPNVTISILELRNKEFKHNNPAHYL 1250  
Db 361 ADGNKLAPVALAVIDGSLVVGDFNYTRIRFPSPNVTISILELRNKEFKHNNPAHYL 420  
Qy 1251 AVDPVSGSLYSDTNSRRIRYKSLSGTKDLAGNSEVAVGTGEQCLPDEARCGDGGKAI 1310  
Db 421 ATDPVTGDLVYSDTNTIRYKSLSGTKDLAGNSEVAVGTGEQCLPDEARCGDGGKAI 480  
Qy 1311 DATLMSPRGIAVDKNGLMYFVDATMIRKVDQNGIISTLLGNSDLTAVRPLSCDSMVAQ 1370  
Db 481 EATLMSPRGMAVDKNGLMYFVDATMIRKVDQNGIISTLLGNSDLTAVRPLSCDSMVAQ 540  
Qy 1371 VRLWPTDPLAVNPMDSLYLVLENVILRITENHQVYSITAGRPMHCQVPGIDISLSKLAIH 1430  
Db 541 VRLWPTDPLAVNPMDSLYLVLENVILRITENHQVYSITAGRPMHCQVPGIDISLSKLAIH 600  
Qy 1431 SALESASAIASHTGVLYITETDEKKINLRQVTTNGEICILLAGAASDCCKNDVNCY 1490  
Db 601 TTLESATAIAVSYSGVLYITETDEKKINLRQVTTNGEICILLAGAASDCCKNDVNCY 660  
Qy 1491 SGDDAYATDAILNSPSSSLAVAPDGTIYIADLGNIRIRAVSKNPKVNLAFNAFNAASPEQ 1550  
Db 661 QSGDGYAKDAKLSAPSSLAASPDGTLXIADLGNIRIRAVSKNPKVNLAFNAFNAASPEQ 720  
Qy 1551 ELYVFNADGHIQYVSVLTGELYLYNFTYSTDNVDTYELTDNNGNSLUKIRRDSGGRHLLM 1610  
Db 1551 ELYVFNADGHIQYVSVLTGELYLYNFTYSTDNVDTYELTDNNGNSLUKIRRDSGGRHLLM 1610

Db 721 ELYIFDINGTHQYTSVLTGCDYLYNFYSNDNDITAVTDSNGNTLIRRDPNRMPVRVVS 780  
QY 1611 PDNQIITLVTGNGKLVSTQNLGLMTYDGNLTGLATKSDQETGWTTFYDHEGRLT 1670  
Db 781 PDNQIITLVTGNGKLVSTQNLGLMTYDGNLTGLATKSDQETGWTTFYDHEGRLT 840  
QY 1671 NVTRPGVVTSLHREMEKSIITDIENSRRDDVTVITNLSSEASVTVVODQVRNSYQLC 1730  
Db 841 NVTFPGVVTNLHGDMDKAITVDIESSEEDVSITNLSIDSFTYVWQDQLRNSYQIG 900  
QY 1731 NNGTLVAVMANGMISFHSSEPHVLACTITPTIGRCNISLPMENGLANSIEWLRKEQIKK 1790  
Db 901 YDGLRIITVAGSDSHYQEPHVLACTAMPTVAKRNWTLPGENGQNLVEMRFKEQAQK 960  
QY 1791 VTFGRKRLVHGRNLSIDYDRIRTEKIIDHRRKFTLRIYDQGRPELWLPSSGLAAV 1850  
Db 961 VNVFGRKRLVHGRNLSIDYDRIRTEKIIDHRRKFTLRIYDQGRPELWLPSSGLMAV 1020  
QY 1851 NVSYFNGRLAGLQRGAMSERTDIDQGRIVSRMFADGKWSYSYLDKSMVLLQSQRQY 1910  
Db 1021 NVTSYTGQIATQRTTSEKVDYDQGRIVSRMFADGKWSYSYLDKSMVLLQSQRQY 1080  
1911 IFEYDSSDRLLAVTPSPVARHSMSTHTSIGYIRNIYNPPESNASVIFDYSDDGRIILKTSF 1970  
Db 1081 IFEYDMDRLSALTSPVARHSMSTHTSIGYIRNIYNPPESNASITDNEGLLLQTA 1140  
QY 1971 LGTGROVFKYKGLSKLSEIVYDSTAVTEGYDETTGVLKMNLSQSGFCTIRYKIGPL 2030  
Db 1141 LGTSRRVLEKRYRQRLSEILYDSTRVSTYDETAGVLTAVNLSQSGFCTIRYKIGPL 1200  
QY 2031 VDKQIYRFSEEGVMNAREDYTHDNSFRIASIKPVISETPLPVDLYRDEISGKVEHFK 2090  
Db 1201 IDQYIFRSEEGVMNAREDYTHDNSFRIASIKPVISETPLPVDLYRDEISGKVEHFK 1259  
QY 2091 FGVIYDIQNIITAVMTLSKHFDTGRIKEVOYEMFRSLMYMTVQYDSMGVIRKREIK 2150  
Db 1260 FGVIYDIQNIITAVMTLSKHFDTGRIKEVOYEMFRSLMYMTVQYDSMGVIRKREIK 1319  
QY 2151 LGPYANTTYDYDGDGQLOQSVAVNDRPTWRYSYDLNGLHLLNPGNSVRLMPLRYDUR 2210  
Db 1320 IGPFAANTTYAYEYVDGQLOQSVAVNDRPTWRYSYDLNGLHLLNPGNSVRLMPLRYDUR 1379  
QY 2211 DRTIRLGDVQYKTDGQYLCQSGSDIFEYNSKGLLTRYNKASGWSVQYRQYDVGRRASY 2270  
Db 1380 DRTIRLGDVQYKTDGQYLCQSGSDIFEYNSKGLLTRYNKASGWSVQYRQYDVGRRASY 1439  
QY 2271 KTNLGHLLQYFYSDLHNPTRITHVYHNSSEITSLYDQGLHFLAMESSGSEYVVASDN 2330  
Db 1440 KTNLGHLLQYFYSDLHNPTRITHVYHNSSEITSLYDQGLHFLAMESSGSEYVVASDN 1499  
2331 TGTPLAVFSLMILKQLOYTAYGEIYDSDNPDQFQWVIGFHGGLYDPLTKLVHFTQRYD 2390  
Db 1500 TGTPLAVFSLMILKQLOYTAYGEIYDSDNPDQFQWVIGFHGGLYDPLTKLVHFTQRYD 1559  
QY 2391 VLAGRWTSPTYTMKVNKGEPAPNLYMFKSNPLSSELDLKNYTDVKSWSLVMFEGFQLS 2450  
Db 1560 VLAGRWTSPTYTMKVNKGEPAPNLYMFKSNPLSSELDLKNYTDVKSWSLVMFEGFQLS 1619  
QY 2451 NIIPGPPRAKMYFVPPPYELSESQASENGOLITGVQOQTERHNOAFMAL---EGQVIRK 2506  
Db 1620 NIIPGPPRAKMYFVPPPYELSESQASENGOLITGVQOQTERHNOAFMAL---EGQVIRK 1676  
QY 2507 KUHASTREKAGH-----WFAATTPPIGKIMFAIREGRVTTGVSSIASDSRKRVASVLNN 2561  
Db 1677 -----SRRRAGGAQSWLWTFATVKSITGKGMVLAVSGRVQTNVLNANEDCICKVAAVLNN 1731  
QY 2562 AVYLDKMHYSIEGKDPHYFVKIGSADGLVLTGTTIGR 2599  
Db 1732 AYLENLHFTIEGKDPHYFVKIGSADGLVLTGTTIGR 1769

RESULT 15  
Q9JLC0

ID Q9JLC0 PRELIMINARY; PRT; 930 AA.  
AC Q9JLC0;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE OD22 (FRAGMENT).  
GN OD22.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ben-Zur T., Feige E., Motro B., Wides R.;  
RT "The mammalian Odc gene family: Homologs of a Drosophila pair rule  
RT gene with expression implying distinct yet overlapping developmental  
RT roles";  
RL Dev. Biol. 117:107-120(2000).  
DR EMBL; AF195419; AAF28317.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 930 AA; 106506 MW; 98879795182369DD CRC64;

Query Match 33.2%; Score 4848; DB 11; Length 930;  
Best Local Similarity 98.9%; Pred. No. 0;  
Matches 920; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

QY 1804 NLLSIDYDRNIRTEKIIDHRRKFTLRIYDQGRPELWLPSSGLAAVNVSYFNGRLAGL 1863  
Db 1 NLLSIDYDRNIRTEKIIDHRRKFTLRIYDQGRPELWLPSSGLAAVNVSYFNGRLAGL 60  
QY 1864 QRGAMSERTDIDQGRIVSRMFADGKWSYSYLDKSMVLLQSQRQYIFEYDSSDRLLAV 1923  
Db 61 QRGAMSERTDIDQGRIVSRMFADGKWSYSYLDKSMVLLQSQRQYIFEYDSSDRLLAV 120  
QY 1924 TWPSPVARHSMSTHTSIGYIRNIYNPPESNASVIFDYSDDGRIILKTSFDTGQVFKYK 1983  
Db 121 TWPSPVARHSMSTHTSIGYIRNIYNPPESNASVIFDYSDDGRIILKTSFDTGQVFKYK 180  
QY 1984 LSKLSIYVYDSTAVTEGYDETTGVLKMNLSQSGFCTIRYKIGPLVDKQIYRSEECM 2043  
Db 181 LSKLSIYVYDSTAVTEGYDETTGVLKMNLSQSGFCTIRYKIGPLVDKQIYRSEECM 240  
QY 2044 VNAREDYTHDNSFRIASIKPVISETPLPVDLYRDEISGKVEHFKGFGVIYDINQIIT 2103  
Db 241 VNAREDYTHDNSFRIASIKPVISETPLPVDLYRDEISGKVEHFKGFGVIYDINQIIT 300  
QY 2104 TAVMTLSKHFDTGRIKEVOYEMFRSLMYMTVQYDSMGVIRKREIKLGPYANTTYTYD 2163  
Db 301 TAVMTLSKHFDTGRIKEVOYEMFRSLMYMTVQYDSMGVIRKREIKLGPYANTTYTYD 360  
QY 2164 YDGDGQLOQSVAVNDRPTWRYSYDLNGLHLLNPGNSVRLMPLRYDLRDRITRLGDVQYKI 2223  
Db 361 YDGDGQLOQSVAVNDRPTWRYSYDLNGLHLLNPGNSVRLMPLRYDLRDRITRLGDVQYKI 420  
QY 2224 DDGGLCQSGSDIFEYNSKGLLTRYNKASGWSVQYRQYDVGRRASYKTNLGHLLQYFVS 2283  
Db 421 DDGGLCQSGSDIFEYNSKGLLTRYNKASGWSVQYRQYDVGRRASYKTNLGHLLQYFVS 480  
QY 2284 DLHNPTRITHVYHNSSEITSLYDQGLHFLAMESSGSEYVVASDNTGTPLAVSINGL 2343  
Db 481 DLHNPTRITHVYHNSSEITSLYDQGLHFLAMESSGSEYVVASDNTGTPLAVSINGL 540  
QY 2344 MIKQLOYTAYGEIYDSDNPDQFQWVIGFHGGLYDPLTKLVHFTQRYDQVYDLAGRWTSPDYTM 2403  
Db 541 MIKQLOYTAYGEIYDSDNPDQFQWVIGFHGGLYDPLTKLVHFTQRYDQVYDLAGRWTSPDYTM 600  
QY 2404 WKNVKEPAPNLYMFKSNPLSSELDLKNYTDVKSWSLVMFEGFQLSNIIIPGFPRAKMYF 2463  
Db 601 WKNVKEPAPNLYMFKSNPLSSELDLKNYTDVKSWSLVMFEGFQLSNIIIPGFPRAKMYF 660  
QY 2464 VPPPYELSESQASENGOLITGVQOQTERHNOAFMALEGOVITTKLHASITREKAGHFAFT 2523  
Db 2523 VPPPYELSESQASENGOLITGVQOQTERHNOAFMALEGOVITTKLHASITREKAGHFAFT 2523







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 11:02:55 ; Search time 63.81 Seconds  
(without alignments)  
4757.321 Million cell updates/sec

Title: US-09-800-198-8  
Perfect score: 14581  
Sequence: 1 MDVKDRRHRSITRGCGKEC.....ELADSSNIQFLRQNMGRK 2733

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

ched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14581	100.0	2733	22 AAU08680	Human FCPR3b polyp
2	14520.5	99.6	2724	22 AAU08681	Human FCPR3f polyp
3	6567	45.0	2136	22 AAU08695	Human protein SEQ
4	3915.5	26.9	2515	22 ABB71354	Drosophila melanog
5	3880.5	26.6	1045	22 ABB92858	Human protein sequ
6	3628.5	24.9	964	22 ABB93294	Human protein sequ
7	2837	19.5	1015	22 AAM79679	Human protein sequ
8	2335	16.0	625	22 AAB92780	Human protein sequ
9	2122.5	14.6	429	22 AAU00392	Human secreted pro
10	2022.5	13.9	746	22 ABB07028	Novel human diagno
11	1931	13.2	381	22 AAU08679	Human FCPR3a polyp

12	1931	13.2	381	22	AAU0861141	Human NOV11 protei
13	1708.5	11.7	402	22	ABG04674	Novel human diagno
14	1250	8.6	242	22	ABG04673	Novel human diagno
15	1205	8.3	865	22	ABG03234	Novel human diagno
16	1203.5	8.3	768	19	AAW44817	Human gamma-heregu
17	1203.5	8.3	768	20	AAU06639	Gamma-hereregulin
18	1203.5	8.3	768	21	AAU06639	Human hereregulin
19	1203.5	8.3	768	22	AAU06639	Human hereregulin
20	1167	8.0	777	22	ABG5598	Drosophila melanog
21	1110	7.6	587	22	ABG5598	Drosophila melanog
22	1043.5	7.2	1124	22	ABG5598	Drosophila melanog
23	726	5.0	1810	17	AAU04563	Chicken cytotactin
24	713	4.9	136	22	ABG04672	Novel human diagno
25	680	4.7	2201	22	AAU06639	Human tenascin-C
26	668	4.6	527	19	AAW44818	Human gamma-heregu
27	666.5	4.6	2199	17	AAU04562	Human cytotactin
28	618	4.2	4618	22	AAU04562	Human polypeptide
29	596.5	4.1	214	22	AAU04562	Human SEC9 protein
30	560	3.8	151	21	AAU04562	Human ORFX ORF181
31	455	3.1	494	21	AAU04562	TIDE protein encoded
32	455	3.1	567	21	AAU04562	protein encoded by
33	450.5	3.1	469	21	AAU04562	Chick Serrate. Ga
34	444	3.0	1193	17	AAU04562	protein encoded by
35	444	3.0	1193	21	AAU04562	Chick Serrate. Ga
36	438.5	3.0	2321	19	AAU04562	Human Notch3 prote
37	434	3.0	636	22	AAU04562	Rat TANGO 272 SEQ
38	430.5	3.0	2471	20	AAU04562	Human Notch2 (hum
39	430	2.9	1010	20	AAU04562	Human JAGGED1 solu
40	430	2.9	1036	18	AAU04562	Proliferation and
41	430	2.9	1187	18	AAU04562	Human JAGGED1 solu
42	430	2.9	1208	19	AAU04562	Human JAGGED1 solu
43	430	2.9	1218	17	AAU04562	Human Serrate-1 (H
44	430	2.9	1218	19	AAU04562	Human Serrate 1.
45	430	2.9	1218	20	AAU04562	Human JAGGED1 prot

ALIGNMENTS

RESULT	1
AAU08680	
ID	AAU08680 standard; Protein; 2733 AA.
AC	AAU08680;
XX	
DT	18-DEC-2001 (first entry)
XX	
DE	Human FCPR3b polypeptide sequence.
XX	
XX	Human; FCPR3; myelogenous leukaemia; carcinoma; melanoma; glioma;
KW	astrocytoma; congenital neonatal alloimmune thrombocytopenia; infection;
KW	neurological disorder; neurodegenerative disorders; nerve trauma;
KW	familial myelodysplastic syndrome; Charcot-Marie-Tooth neuropathy;
KW	demyelinating disorder; familial myelodysplastic syndrome;
KW	mental health condition; immunological disorder; allergy; infertility;
KW	bronchial asthma; Avellino type eosinophilia; lung disease; deafness;
KW	reproductive disorder; reproductive disorder; glycoprotein Ia deficiency;
KW	desmoid disease; turcot syndrome; liver cirrhosis; hepatitis C; virucide;
KW	gastric disorders; pancreatic disease; Schistosoma mansoni infection;
KW	Spinocerebellar ataxia; Plasmidum faiciparum parasitaemia; diabetes;
KW	Corneal dystrophy-Greenough type I; Corneal dystrophy-lattice type I;
KW	Reis-Bucklers corneal dystrophy; cytostatic; immunosuppressive;
KW	antidiabetic; antidiabetic; antidiabetic; antidiabetic; antidiabetic;
KW	gynaecological; antidiabetic; antidiabetic; antidiabetic; antidiabetic;
XX	gene therapy; FCPR3b; neutrestin-like protein.
OS	Homo sapiens.
XX	
PN	WO200166747-A2.
XX	
PD	13-SEP-2001.
XX	

PF 05-MAR-2001; 2001WO-US07160.  
XX  
PR 03-MAR-2000; 2000US-186592P.  
PR 03-MAR-2000; 2000US-186718P.  
PR 06-MAR-2000; 2000US-187293P.  
PR 06-MAR-2000; 2000US-187294P.  
PR 17-MAR-2000; 2000US-190400P.  
PR 07-APR-2000; 2000US-196018P.  
PR 03-JAN-2001; 2001US-259548P.  
XX (CURA-) CURAGEN CORP.  
XX  
XX Vernet CAM, Fernandes E, Shimkets RA, Herrmann JL, Majumder K;  
PI MacDougall J, Mishra V, Mezes PS, Rastelli L;  
XX  
XX WPI: 2001-596837/67.  
DR N-FSDB: AAS14085.  
XX  
XX Novel polypeptides designated as FCTR polypeptides, useful in  
PT detection, prevention and treatment of a broad range of pathological  
PT states -  
XX  
XX Claim 1; Page 35-36; 215pp; English.  
XX  
XX The invention relates to human FCTR polypeptides, FCTR1-FCTR7, and the  
CC nucleic acids encoding them. These sequences are useful for the treatment  
CC or prevention of numerous disorders including myelogenous leukaemia,  
CC carcinoma, melanomas, gliomas, astrocytomas, congenital neonatal  
CC alloimmune thrombocytopenia, neurological disorders, neurodegenerative  
CC disorders, nerve trauma, familial myelodysplastic syndrome,  
CC Charcot-Marie-Tooth neuropathy, demyelinating Gardner syndrome, familial  
CC myelodysplastic syndrome, mental health conditions, immunological  
CC disorders, allergy and infection, bronchial asthma, Avelino type  
CC eosinophilia, lung diseases, reproductive disorders, infertility, male  
CC and female reproductive disorders, deafness, glycoprotein Ia deficiency,  
CC desmoid disease, turcot syndrome, liver cirrhosis, hepatitis C, gastric  
CC disorders, pancreatic diseases such as diabetes, Schistosoma mansoni  
CC infection, spinocerebellar ataxia, Plasmodium falciparum parasitaemia,  
CC Corneal dystrophy-Greennow type I, Corneal dystrophy-lattice type I and  
CC Reis-Bucklers corneal dystrophy. This sequence represents FCTR3b, a  
CC neurestin-like protein.  
XX  
SQ Sequence 2733 AA;

Query Match 100.0%; Score 14581; DB 22; Length 2733;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2733; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MDVKRRHRLTRGRCGKECTSSLDSEDCRVPTQKSYSSSETLKAYDHRMHNHGNR 60  
|||||  
1 mdvkdrrhrsltrgrcgkectssldsedcrrvptqkssysselkaydhrmhygnr 60  
|||||

61 VTDLTHRESDPEPGTNTFTLAELGICPPSHRSGYCSMDGTLHGYSLSLSTGSDADSDE 120  
|||||  
61 vcdllhresdeprgntntftlaelgicpsphrsygcsmgllhgysslstgsdadsde 120  
|||||

121 GMSPEHAIRLWGRGIKRRSSGLSRNSALTLDSDNENKSDDENGRPTPTSSPLL 180  
|||||  
121 gmspehairlwgrgikrrssglsrnsaltldsdnenksddengrptptsspsll 180  
|||||

181 PSAQLPSSHNPPVSCQPLDLSNTSHQIMDTNPDEEFPSPNYSLLRACSGPQASSGPP 240  
|||||  
181 psaqlpsshnppvscqplldlsntshqimdtnpdeefspnysyllracsgpqassgpp 240  
|||||

241 NHHSSOSTLRPLPPPHNHTLSHHSSANSLNRSNTNRSSQTHAPAPNDLATTPESVQ 300  
|||||  
241 nhhsostlrplppphnhtlshhssanslnrsntnrssqthapapndlattpesvq 300  
|||||

301 LODSWLNSNVLPRHFLFKTSSGSTPLFSSSSPGYPLTSGTVYTPPPRLPRNTFSRK 360  
|||||  
301 lqdswwlnsnvnlprhflfkstsgstplfsssspgypltsgtvytppprllprntfsrk 360  
|||||

Qy 361 AFKLKPKSKYCSWKCAALSAIAAALLAILLAYFIVPWSLKNSSIDSGEAEVGRVTVQEV 420  
|||||  
Db 361 afklkpkskycswkcaalaaiaaallaillayfivpwslnssidsgeaevgrvtqev 420  
|||||

Qy 421 PPGVFWRSQIHIISQPOFLKFNISLGKDALFGVYIRRGILPPSHAQYDFMERLDGKESVW 480  
|||||  
Db 421 ppgvfwrsqihisqpqlkfnislgkdalfgvyirrgilppshaqydfmerldgkekswv 480  
|||||

Qy 481 ESPRERRSIQTLVQNEAVFVOYLDVGLWHLAFYNDGKDEKEMVSFNTVVLDSVQDCPRNCH 540  
|||||  
Db 481 esprerrsiqtlvqneavfvoyldvglwhlafyndgkdekemvsfntvvlsvdqpcprnch 540  
|||||

Qy 541 NGECVSGVCHCFPFGLGADCAKAAACPVLCSGNGGYSGKTCOCYSGWKGAECEDVPMNOCI 600  
|||||  
Db 541 ngecvsgvchcfpgfigadcaakaacpvlcsngngysgkctccysgwkgaecdvpnmnci 600  
|||||

Qy 601 DPSCGGHSCIDGNCVCSAGYKGEHCEVEDCLDPTCSSHGVGVNAGECLCSPGWGLNCEL 660  
|||||  
Db 601 dpscgghscidgncvcvcsagykghehcevedcldpctcsshgvnagelcspwgglncel 660  
|||||

Qy 661 ARVOCPCQCSGHGTLPDTGLCSDPNMMGPDSCSVEVCSVDCGTHGVGIGACRCEEGWT 720  
|||||  
Db 661 arvocpcqcsghgtltpdtglcscdpnmwmpdcsvcvcsdcgthgvigacrcceegwt 720  
|||||

Qy 721 GAACDQRVCHPRCIEHGTCKDGKCECREGWNGEHCITIGRTAGTETDCDPLCNGNGRGT 780  
|||||  
Db 721 gaacdqrvcprciehgtckdgkcecregwngehctigrtagtetdcdplcngngret 780  
|||||

Qy 781 LGOQSWQCVCOTGWRGPCCNVAMETSCADNKDNEGDLVDCLDPCCLQSAACONSLLCRG 840  
|||||  
Db 781 lqoswqcvcotgwrpgpcnvmetscadnkdnegdlvdcldpccclqsacqnsllcrg 840  
|||||

Qy 841 SRDPLDIIIOQGTOWPAVKSFYDRILKLAGKDSHTIIPGENPFNSLSVLIRGQVVTG 900  
|||||  
Db 841 srdpldiioqgtowpavksfydrilklagkdsthiipgenpfnsllsvlirgqvvtg 900  
|||||

Qy 901 TPLVCVNVSVFKYPKYGTITRQDCTFDLIANGASLTLHFERAPFMSQERTVLPWNSF 960  
|||||  
Db 901 tplvcvvnsvfkykgyttrdqctfdliangaslthferapfmsqertvlpwnsf 960  
|||||

Qy 961 YAMDTLVNKTENSIPTSCDLSGFVRPDPITIISSPLSTFFSAAGQNPVPTQVLHEBIE 1020  
|||||  
Db 961 yamdtlvnmkteensiptscdlsghfvrpdpitiiissplstffsaagqnpvptqlhebie 1020  
|||||

Qy 1021 LPSGNWKLRYLSSRTAGYKSLKTKITWTOSTVPLNIRVHLMVAVEGHLFQKSFQASPLA 1080  
|||||  
Db 1021 lpgsnwklrylssrtagysllktkitwtostvplnlirvhlmvaveghlfqksfqaspla 1080  
|||||

Qy 1081 STFIWDKTDAYGQRYVGLSDAVSVGFYETCPSLILWEKRTALLQGFELDPSPNLGGWSL 1140  
|||||  
Db 1081 stfiwdktdaygqrvyglsdavsvgfyetcpslilwekrtallqgfeldpnsnlggwsl 1140  
|||||

Qy 1141 DKHHLNVKSGILHKGTEGNOFLTQOPAIITISIMGNRRRSISPCSNGLAEGNKLAPV 1200  
|||||  
Db 1141 dkhhlnvksgilhkgtegenqflttqpaititsimgnrrrsiscpscnглаegnklapv 1200  
|||||

Qy 1201 ALAVGIDSLVGDENVYIRRPFPSPNRVTSILELRNKEFHNNPAKYYLAVDPVSGSLY 1260  
|||||  
Db 1201 alavgidslvgdenvyirrfpfpnrvtlsilelrnkefhnnpakyylavdpvsgsly 1260  
|||||

Qy 1261 VSDTNSRRIYRVKSLSGTKDLAGNSEVAVGTCEQCLPDEARCGDGGKAIDATLMSPRGI 1320  
|||||  
Db 1261 vsdtnsrriyrvkslsgtkdlagnsevvagtceqclpdearcgddggaiddatlmprgi 1320  
|||||

Qy 1321 AVDKNGLMYFVDATMIRKVDQNGIISTLLGSNDLTAVRPLSCDSDMDVAQVRLEWPTDLA 1380  
|||||  
Db 1321 avdknglmvfvdattmrkvdqngiistllgsndltavrplescdsdmdvaqvrlewptdla 1380  
|||||

Qy 1381 VNPMDNSLYLVENNVLIRITENHQVSIITAGRPMHCQVPGIDYSLSKLAHSALESASAIA 1440  
|||||  
Db 1381 vnpmdnslylvennvliritenhqvsiitagrpmhcqvpgidyslsklaihalesasala 1440  
|||||

Qy 1441 ISHTGVLYITETDEKKINRLRQVTNNGEITCLLAGAASDCDCKNDVNCYSGDDAYATDA 1500  
|||||

Db 1441 ishtgviyifetdekklnrlrqvtntngeicllagaaedcdckndvncncysgddaytada 1500  
QY 1501 ILSPSSLAVAPDGTIYIADLGNIRIRAVSKNKPVLNAFYEAASPGCELYEVFNADGI 1560  
Db 1501 ilspsslavapdgltiyadiagnlriravsknkpvlinafqyeaaspgdeqelyyfnadgi 1560  
QY 1561 HQTVSLVTGEYLYNFYSTDNDVTELIDNNGNSLKIIRDSSGMPRHLLPMDNQIITLV 1620  
Db 1561 hqtvslvtgeylynfystdndvtelidnngnslkrrdssgmprhlmpdnqiliclv 1620  
QY 1621 GTNGGLKVVSTONLEGLMYDGNLTGATKSDETGHTTFYDHDHGRNLNVRPTGVVT 1680  
Db 1621 gtngglkvstqnlleqlmtydngntllatksetgtwttydydhegrltnvtrptgvvt 1680  
QY 1681 SLHREMEKSTIDIENSRRDDVTITNLSSVEASYTVVODQVRNSYOLCNGNGLTRVMYA 1740  
Db 1681 slhremekstidlenenrddvtitnlssveasytvdqdvznsyqlcngnqlrvmya 1740  
Db 1741 NGMISPHSEPHVLAGTITPTIGRCNLSLPMENGLNSIEWRLRKEQIKGVITFGRKLRY 1800  
Db 1741 ngmishsephvlagtitptigrncnslpmenglnsiewrlrkeqikgvitfgrklrv 1800  
QY 1801 HGRNLLSIDYDRNIRTEKIYDDHRKFTLRIYDOVGPRPFLWLPSSGLAANVSYFFNGRL 1860  
Db 1801 hgrnllsdydrnirtekiyddhrkftlriydgvrpflwlpssglaavnvsvffngrl 1860  
QY 1861 AGLQGAAMSERTIDKGRIVSRMFADGKVMYSYLDKSMVLLLOSQRQYIFEVDSDDL 1920  
Db 1861 aglqgamsertidkgrivsrmfadgkvmysyldksmvlllqsqrqyifeydsddl 1920  
QY 1921 LAVTMPVARHSTHTSISYIRNIYNPPESNASVIFDYDDGRILKTSFLGTGRQVYFK 1980  
Db 1921 lavtmpvarhsthtsisyirniynppesnasvifdydsdgrilktstflgtgrqvyfk 1980  
QY 1981 YGKLSKLSIYVYSTAVTFGDETTGVLMKNVLOSQGFSTIRYRKIGPLVDQIYRFE 2040  
Db 1981 ygklsklsiyvystavtfgedttgvlmknvlosqgfstiryrkigplvdqiyrfe 2040  
QY 2041 EGMVNARFDTYTHDNSPRIASIKPVISETPLPVDLYRYDEISGRVHFGRFGVYIYDQ 2100  
Db 2041 egmvnarfdtythdnspriasikpvi setplpvdlyrydeisgrvfhgrfgvlyydaq 2100  
QY 2101 IITAVMTLSKHFDTHGRKEVQYEMFSLMYMTVOYDSMGRVIRKRELKLGYPYANTTKY 2160  
Db 2101 iitavmtlskhfdthgrikevqyemfslmymtvqydsmgrvirkrelklgpyanttky 2160  
Db 2161 TYDYGQGLQSVAVNDRPTWRYSYDLNGLNLLNPGNSVRLMPLRYDLRDRITRLGDVQ 2220  
Db 2161 tydydgqglqsvavndrptwrysdydingnlhllnpgnsvrlmplrydlrdrtrldgvdq 2220  
QY 2221 YKIDDDGLQCGSDIFEYNSKGLLTRYANKASGWSYQYRYDGVGRASVYKTLNGLHLLQY 2280  
Db 2221 ykiddgylqcgdsdifeynskgltryankasgwsyqrydygvgrasvktlghllqy 2280  
QY 2281 FYSOLHNPTRITHYVNSHNSITSLYDLOGLHFAMESSSGEEYVYASDNTGTPLAYFSI 2340  
Db 2281 fysolhnptritthyvnhnsaitslydloqlhfamesssgeeyvvasdntgtplavfsi 2340  
QY 2341 NGLMIKQLOYTAYGEIYYDSNPQFQWIGVPHGGLYDPLTKLVHFTQRDQDVLAGRWTSPPD 2400  
Db 2341 nglmikloytaygeiyydsnpdqfwigvphgglydpltklvhftqrdqdvlagrwtsppd 2400  
QY 2401 YTMKNVKGKPPAPNLYMFKSNPNLSSELDELKNVYTDVQKSWLMVFMFGQLSNIIPGFPRAK 2460  
Db 2401 ytmknvkgkppapnlymfksnppnlsseldelknyvtdvqkswlmvmfgqlsnlpgfpfrak 2460  
QY 2461 MYFVPPPYELSESQASENGQILITGVQQTTERHQAFAFMALEQGVITKKLHASIREKAGHWF 2520  
Db 2461 myfvpypyelssesqasengqiligtvqqtterhqaafmaleqgvitkklhasirekaghwf 2520  
QY 2521 ATTPPIITGKIMFAIKGRVTTGVSSSTASEDSRKAASVLLNAYYLDKMHYSIEGKQTHYF 2580  
Db 2521 attppiitgkimfaiagrvtgvtgssstasedsrkavllnayyldkmhysieqkthyf 2580

Db 2521 atttligkimfaikegrvtgvtgssstasedsrkavlnnayyldkmhysieqkthyf 2580  
QY 2581 VKIGSADGDVLTGTTIGRKVLESGVNVTVSQPTLLVNGTRRFTNTEFOYSTLLLSIRY 2640  
Db 2581 vkigsadgdvltgttigrkvlesgvnvtvsqptllvngtrtrfnlefyostlllsiry 2640  
QY 2641 GLTPDTLDEKARVLDQARALGTAWAKEQKARDCREGSRLWTEGEKQOOLJSTGRVQG 2700  
Db 2641 gltpdtldeekarvldqaralgtawakeqkardregsrllwtegeqkqlstgrvqg 2700  
QY 2701 YEGYVLPVQYEPYELADSSSNIOFLRONEMGKR 2733  
Db 2701 yegyvlpvqyepeladsssnlqflrqnemgkr 2733  
RESULT 2  
AAU08681  
ID AAU08681 standard; Protein; 2724 AA.  
XX AC AAU08681;  
XX AC  
DT 18-DEC-2001 (first entry)  
XX DE Human FCTR3f polypeptide sequence.  
XX KW Human; FCTRX; myelogenous leukaemia; carcinoma; melanoma; glioma;  
KW astrocytoma; congenital neonatal alloimmune thrombocytopaenia; infection;  
KW neurological disorder; neurodegenerative disorders; nerve trauma;  
KW familial myelodysplastic syndrome; Charcot-Marie-Tooth neuropathy;  
KW demyelinating Gardner syndrome; familial myelodysplastic syndrome;  
KW mental health condition; immunological disorder; allergy; infertility;  
KW bronchial asthma; Avellino type eosinophilia; lung disease; deafness;  
KW reproductive disorder; reproductive disorder; glycoprotein Ia deficiency;  
KW desmoid disease; turot syndrome; liver cirrhosis; hepatitis C; virucide;  
KW gastric disorders; pancreatic disease; Schistosoma mansoni infection;  
KW Spinocerebellar ataxia; Plasmidum falciparum parasitaemia; diabetes;  
KW Corneal dystrophy-Greenow type I; Corneal dystrophy-lattice type I;  
KW Reis-Bucklers corneal dystrophy; cytostatic; immunosuppressive;  
KW anti-allergic; antilasthmatic; antifertility; antiinflammatory;  
KW antidiabetic; protozoacide; hepatotropic; virucide; ophthalmological;  
KW gynaecological; antifertility; immunostimulant; auditory; haemostatic;  
KW gene therapy; FCTR3a; neurestin-like protein; FCTR3f.  
XX OS Homo sapiens.  
XX PN WO200166747-A2.  
XX PP 13-SEP-2001.  
XX PF 05-MAR-2001; 2001WO-US07160.  
XX PR 03-MAR-2000; 2000US-186592P.  
PR 03-MAR-2000; 2000US-186718P.  
PR 06-MAR-2000; 2000US-187293P.  
PR 06-MAR-2000; 2000US-187294P.  
PR 17-MAR-2000; 2000US-190400P.  
PR 07-APR-2000; 2000US-196018P.  
PR 03-JAN-2001; 2001US-259548P.  
XX PA (CURA-) CURAGEN CORP.  
XX PA Vernet CAM, Fernandes E, Shimkets RA, Herrmann JL, Majumder K;  
PI Macdougall J, Mishra V, Mezes PS, Rastelli L;  
XX XX  
XX WPI; 2001-596837/67.  
DR N-PSDB; AAS14089.  
XX XX  
PT Novel polypeptides designated as FCTRX polypeptides, useful in  
PT detection, prevention and treatment of a broad range of pathological  
XX states -  
XX Claim 1; Page 39; 215pp; English.  
XX XX

CC The invention relates to human FCTR3 polypeptides, FCTR1-FCTR7, and the  
CC nucleic acids encoding them. These sequences are useful for the treatment  
CC or prevention of numerous disorders including myelogenous leukaemia,  
CC carcinomas, melanomas, gliomas, astrocytomas, congenital neonatal  
CC autoimmune thrombocytopaenia, neurological disorders, neurodegenerative  
CC disorders, nerve trauma, familial myelodysplastic syndrome,  
CC Charcot-Marie-Tooth neuropathy, demyelinating Gardner syndrome, familial  
CC myelodysplastic syndrome, mental health conditions, immunological  
CC disorders, allergy and infection, bronchial asthma, Avellino type  
CC eosinophilia, lung diseases, reproductive disorders, infertility, male  
CC and female reproductive disorders, deafness, glycoprotein Ia deficiency,  
CC desmoid disease, turtoc syndrome, liver cirrhosis, hepatitis C, gastric  
CC disorders, pancreatic diseases such as diabetes, Schistosoma mansoni  
CC infection Spinocerebellar ataxia, Plasmodium falciparum parasitaemia,  
CC Corneal dystrophy-Greenow type I, Corneal dystrophy-lattice type I and  
CC Reis-Bucklers corneal dystrophy. This sequence represents FCTR3f, a  
CC homologue of FCTR3a protein.  
XX  
SQ Sequence 2724 AA;

Query Match 99.6%; Score 14520.5; DB 22; Length 2724;  
Fast Local Similarity 99.7%; Pred. No. 0;  
Matches 2724; Conservative 0; Mismatches 0; Indels 9; Gaps 1;

QY 1 MDVKRRHRLRGRGRCRVTSSLSDEDCRVTQKSYSSSETLKAYDHDPSRMHYGNR 60  
DB 1 MDVKRRHRLRGRGRCRVTSSLSDEDCRVTQKSYSSSETLKAYDHDPSRMHYGNR 60  
QY 61 VDTLTHRESDEPRGNTFTLAELGICPSPHRSGYCDPMGLHOGYSLSGTGSDADSDE 120  
DB 61 VDTLTHRESDEPRGNTFTLAELGICPSPHRSGYCDPMGLHOGYSLSGTGSDADSDE 120  
QY 121 GKMSPHAIIRLGRGTRKRRSGLSSRENSALTLDSDNENKSDDENGRPIPTSPSLL 180  
DB 121 GKMSPHAIIRLGRGTRKRRSGLSSRENSALTLDSDNENKSDDENGRPIPTSPSLL 180  
QY 181 PSAQLPSSNPPVSCQPLDLSNTSHQIMTNDNPEERSPNSYLLRACSGPOQASSGPP 240  
DB 181 PSAQLPSSNPPVSCQPLDLSNTSHQIMTNDNPEERSPNSYLLRACSGPOQASSGPP 240  
QY 241 NHHSGSTLRPLPPPHNHTLSHHSSANSLSNLSLTNRSSQTHAPAPNDLATTPESVQ 300  
DB 241 NHHSGSTLRPLPPPHNHTLSHHSSANSLSNLSLTNRSSQTHAPAPNDLATTPESVQ 300  
QY 301 LODSWLNSNPLETRHFLFKTSGSTPLFSSSPGYPLTSGTVTPPPRLPLRNTFSRK 360  
DB 301 LODSWLNSNPLETRHFLFKTSGSTPLFSSSPGYPLTSGTVTPPPRLPLRNTFSRK 360  
QY 361 AFKLLKPSKYCSWKCAALSAIAAALLAILLAYFVPSLKNSSSDSGEAEVGRVTQEV 420  
DB 361 AFKLLKPSKYCSWKCAALSAIAAALLAILLAYFVPSLKNSSSDSGEAEVGRVTQEV 420  
QY 421 PPGVFWRSQIHISQPFKFNLSLCKDALFGVYIRRGLPSPSHAQYDEMERLDGKEKWSV 480  
DB 421 PPGVFWRSQIHISQPFKFNLSLCKDALFGVYIRRGLPSPSHAQYDEMERLDGKEKWSV 480  
QY 481 ESPRRRSITQLVQNEAVFVQYLDVGLMHLAFYNDGKEMVSFNTVLDSDQDCPRNCH 540  
DB 481 ESPRRRSITQLVQNEAVFVQYLDVGLMHLAFYNDGKEMVSFNTVLDSDQDCPRNCH 540  
QY 541 GNGECVSGVCHCFPGFLGADCAKACPVLCSSNGQYSGTKOCYSGWKGAECDVPMNQCI 600  
DB 541 GNGECVSGVCHCFPGFLGADCAKACPVLCSSNGQYSGTKOCYSGWKGAECDVPMNQCI 600  
QY 601 DPSCGGHSCIDGNCVCSAGYKGEHCEEVDCUDPTCSSHGVCSVNGECLCSPGMGLNCEL 660  
DB 601 DPSCGGHSCIDGNCVCSAGYKGEHCEEVDCUDPTCSSHGVCSVNGECLCSPGMGLNCEL 660  
QY 661 ARVQCPDQCSGHGTLYPDTGLSCSDPNMNGPDCSVCEVCSVDCGTHGVCIGGACRCEGWT 720  
DB 661 ARVQCPDQCSGHGTLYPDTGLSCSDPNMNGPDCSVCEVCSVDCGTHGVCIGGACRCEGWT 720

QY 721 GAACDQRVCHPRCIRIEHGTCKDKGCECREGNGEHCITGQTAGTETDCCPDLCNGNRCR 780  
DB 721 GAACDQRVCHPRCIRIEHGTCKDKGCECREGNGEHCITGQTAGTETDCCPDLCNGNRCR 780  
QY 781 LGONSWQOCVQGWGRGPGCNVAMETSCADKNDEGDLVDCLDPPCCCLQSACONSLLCRG 840  
DB 772 LGONSWQOCVQGWGRGPGCNVAMETSCADKNDEGDLVDCLDPPCCCLQSACONSLLCRG 840  
QY 841 SRDPLDIITQQGQTDPAVKSFYDRIKLAGKDSHTIIPGENPFNSLSVSLIRGQVVTIDG 900  
DB 832 SRDPLDIITQQGQTDPAVKSFYDRIKLAGKDSHTIIPGENPFNSLSVSLIRGQVVTIDG 900  
QY 901 TELVGNVSVFYKPYKYGTITRQDQTFDIANGASLTUHFERAPFMSQERTVWLPWNSF 960  
DB 892 TELVGNVSVFYKPYKYGTITRQDQTFDIANGASLTUHFERAPFMSQERTVWLPWNSF 960  
QY 961 YAMDTLVKMTTEENSIPSCDLSGFVRPDIITSSPLSTFESAAGNPITVPTQVLEHETE 1020  
DB 952 YAMDTLVKMTTEENSIPSCDLSGFVRPDIITSSPLSTFESAAGNPITVPTQVLEHETE 1020  
QY 1021 LFGSNVKLYRSSRTAGYKSLIKITMTQSTVPLNLRVHLMVAVEGHLFOKSFQASPNLA 1080  
DB 1012 LFGSNVKLYRSSRTAGYKSLIKITMTQSTVPLNLRVHLMVAVEGHLFOKSFQASPNLA 1080  
QY 1081 STFIWKTDAYQORYGLSDAVVSVGFYETCPSPILMEKRTALLQGFELDSPNLGGWSL 1140  
DB 1072 STFIWKTDAYQORYGLSDAVVSVGFYETCPSPILMEKRTALLQGFELDSPNLGGWSL 1140  
QY 1141 DKHHILNVKSGTLHKGTTGENQFLTQOPAITISIMNGRRRSISCPSCNGLAGNKLLAPV 1200  
DB 1132 DKHHILNVKSGTLHKGTTGENQFLTQOPAITISIMNGRRRSISCPSCNGLAGNKLLAPV 1200  
QY 1201 ALAVGIDSLYVGDFNYIRRIIPSRNVTSLILRNKEFKHNNPAHKYYLAVDPVSGSLY 1260  
DB 1192 ALAVGIDSLYVGDFNYIRRIIPSRNVTSLILRNKEFKHNNPAHKYYLAVDPVSGSLY 1260  
QY 1261 VSDTSNRRIYRVKSLSGTKDLAGNSEVAGTCEQCLPDEARCGDGGKAIDATLMSPRGI 1320  
DB 1252 VSDTSNRRIYRVKSLSGTKDLAGNSEVAGTCEQCLPDEARCGDGGKAIDATLMSPRGI 1320  
QY 1321 AVDKNGLMYFVDATMIRKVDQNGIISTILGNSNDLTAVRPLSCDSSMDVAQVLEWPTDIA 1380  
DB 1312 AVDKNGLMYFVDATMIRKVDQNGIISTILGNSNDLTAVRPLSCDSSMDVAQVLEWPTDIA 1380  
QY 1381 VNPMDNSLYVLENNVILRITENHQSIIAGRPWHQVPGIDYSLSKLAHSALESASATA 1440  
DB 1372 VNPMDNSLYVLENNVILRITENHQSIIAGRPWHQVPGIDYSLSKLAHSALESASATA 1440  
QY 1441 ISHTGVLYTETDEKKINRLROVTTNGETCLLAGAASDCCKNDKNCYSGDDAYATDA 1500  
DB 1432 ISHTGVLYTETDEKKINRLROVTTNGETCLLAGAASDCCKNDKNCYSGDDAYATDA 1500  
QY 1501 ILNPSSSLAVAPDGTIYIADLGNIRAVSKNPNVNAFNOYEAASPGOEYLVFNADGI 1560  
DB 1492 ILNPSSSLAVAPDGTIYIADLGNIRAVSKNPNVNAFNOYEAASPGOEYLVFNADGI 1560  
QY 1561 HOYTVSLVTGEYLYNFTYSTDNDVTLEIDNNGNSLKIRRDSSGMPRHLLMPDNQIITLV 1620  
DB 1552 HOYTVSLVTGEYLYNFTYSTDNDVTLEIDNNGNSLKIRRDSSGMPRHLLMPDNQIITLV 1620  
QY 1621 GTNGLGVVSTQNLELGLMTYDGTNGLLATKSDGTWTFDYDHEGRLTNVTRTGVVT 1680  
DB 1612 GTNGLGVVSTQNLELGLMTYDGTNGLLATKSDGTWTFDYDHEGRLTNVTRTGVVT 1680  
QY 1681 SLHREMKSITIDIENSRRDDVTITNLSSVEASTVYVQDVNRNSYQLCNGNGLRVMYA 1740  
DB 1672 SLHREMKSITIDIENSRRDDVTITNLSSVEASTVYVQDVNRNSYQLCNGNGLRVMYA 1740  
QY 1741 NGMGISFHPSEPHVLGATTIPTIGRCNLSLPMENGLNSLEWRLEKQIKGVITFGRKLVR 1800  
DB 1732 NGMGISFHPSEPHVLGATTIPTIGRCNLSLPMENGLNSLEWRLEKQIKGVITFGRKLVR 1800  
QY 1801 HGRNLLSDYDRNIRTEKIYDDHRRKFTLRIIYDQVGRFPLMLPSSGLAANVSYFFNGRL 1860









Db 1635 wkvgvqaysforgrlneikygdgstmvyafkdmfmglsplkvttrrsdyllgyddag 1694  
Qy 1919 RLAVTMSVARHSMTHTSIGVIRIYNPPESNASVIFDYDDGRILKTSFLGTGRQVF 1978  
Db 1695 alqslctprghhafsqtsgifkyqyyspnrhpfellyndeqdlakihphsgskva 1754  
Qy 1979 KYGKLSKLSIYDSTAVTFYDDETVGLKMNQSGGFSCTIRYR-KIGPLVDKQIYR 2037  
Db 1755 fvhdtagrietlilagslchtyqdttsivksveveqepgfelrrefkyhagilkdeklrf 1814  
Qy 2038 FSEGMVNAREFDYTHDNSFRASIKPVISEPTPLVDLYRYDEISGVKVEHFGKFGVIYD 2097  
Db 1815 gsknsiasarykay-dgnarlsigemaiddkelptrkyksqnlgle-----vvqd 1866  
Qy 2098 I---NOIITTAVMTLKSHF-----DTHGRKEQYEMERSLWMTVOYDSMGRIKVR 2147  
Db 1867 ikltanartvqdsakffaiydydqhgrvksvlmnnvknidvfrleidydlrnriksq 1926  
Qy 2148 ELKLGVPANTKYTYDYDGGQLQSVAVNDRTWRYSDLNGN-LHLLNPGNSVRLMPLR 2206  
Qy 1927 kttfgrstafdklnyadg---hvvevlgtnwnkylfdngntvgvvdggekfn---lg 1979  
Qy 2207 YDLRDRITRLGDVQY-KIDDDGYLCORGSDIFEYNSKGLLTRAYNKA--SGNSVQYRYDG 2263  
Db 1980 yidgrvkvqdvfnnydargfvvkrgeqkyrnnrgqlihsferfqs---yydd 2036  
Qy 2264 VGRASAKTNLGHQLQYFYSDHLNPTRIYVNSHNSITSLSLYDLQGLHFAWESSGEE 2323  
Db 2037 rarlvaahdnknttgqyyanprphlvthvhpksrtcmkifyddrdmlalehed-qr 2095  
Qy 2324 YVASDNTCTPLAVSINGLMTKLOLYTAYGEIYDSDNPDQWVGFHGLYDPLTKLVH 2383  
Db 2096 yvvatdqnsplaffdqngsilvkmkrtfgrliikdkpeffvpidfghglidphtklvy 2155  
Qy 2384 FTQDYDVLGRWTSPTYTMKNGKE---PAPFNLYMFKSNPLSSELDLKNYVTDVKS 2440  
Db 2156 teqrqydvphvgwmpc---lwetlatemshpdtvfyrynmndpnpnkp-qnymldids 2211  
Qy 2441 WLVMFGQLSNIPGPPRAKMYFVPPPYELSESQASENGQITGVQOQTERINQAFM--- 2497  
Db 2212 wlqlfygdlmngssrytklqytpqaslkstlapdfg-vlsgeclvektseksfdd 2270  
Qy 2498 -----ALEQOVITKLHSAIRKAGHWFATTPITGKGIMFAIKGR-----VTTGVSSSTA 2548  
Db 2271 fvpkpllkmpkmrnlrvsvrrg-----vfigevllsrigralvsvvdgsnv 2322  
Qy 2549 SEDSRKVASVLNAYYLDKMHYSIEGKDPHYFV-----KIGSADGDLVLTGTIGRKVLE 2603  
Db 2323 qd---vsvsvfnnsyfld-lhslhdqdvfyfkdnnvklrddneelrrlgmfaisthe 2378  
Qy 2604 SGVNVTVSOTPLLVNGRTRRTFNIEFQVSTLLSIRYGLTPDTLDEKARVLDQARQAL 2663  
Db 2379 isdhggsaakeirlhg-----pdavvlikygvdv---egethrilkhakhtav 2423  
Qy 2664 GTAWAKEQOKARDREGSLWTEGEKQQLLSGRVQVGEYVVLVPEQYPELADSSNIQ 2723  
Db 2424 eraweleklvgaagfggrgdvteeekeelivqhdvgnwngldihshkypqladpgnva 2483  
Qy 2724 FLRQNMGRK 2733  
Db 2484 fqdakrkr 2493  
RESULT 5  
ID AAB92858  
AA92858 standard; Protein; 1045 AA.  
XX  
AC AAB92858;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human protein sequence SEQ ID NO:11431.

XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX Homo sapiens.  
XX EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-0116126.  
XX 29-JUL-1999; 99JP-0248036.  
XX 27-AUG-1999; 99JP-0300253.  
XX 11-JAN-2000; 2000JP-0118776.  
XX 02-MAY-2000; 2000JP-0183767.  
XX 09-JUN-2000; 2000JP-0241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
XX full-length cDNAs defined in the specification, and for the detection  
XX and/or diagnosis of the abnormality of the proteins encoded by the  
XX full-length cDNAs.  
XX Claim 8; SEQ ID 11431; 2537pp + CD ROM; English.  
XX The present invention describes primer sets for synthesizing 5602  
XX full-length cDNAs defined in the specification. Where a primer set  
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
XX to the complementary strand of a polynucleotide which comprises one of  
XX the 5602 nucleotide sequences defined in the specification, where the  
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
XX of an oligonucleotide comprising a sequence complementary to the  
XX complementary strand of a polynucleotide which comprises a 5'-end  
XX sequence and an oligonucleotide comprising a sequence complementary to a  
XX polynucleotide which comprises a 3'-end sequence, where the  
XX oligonucleotide comprises at least 15 nucleotides and the combination of  
XX the 5'-end sequence/3'-end sequence is selected from those defined in  
XX the specification. The primer sets can be used in antisense therapy and  
XX in gene therapy. The primers are useful for synthesizing polynucleotides,  
XX particularly full-length cDNAs. The primers are also useful for the  
XX detection and/or diagnosis of the abnormality of the proteins encoded by  
XX the full-length cDNAs. The primers allow obtaining of the full-length  
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
XX AAH13633 to AAH18742 represent human cDNA sequences; AA92446 to  
XX AA95893 represent human amino acid sequences; and AAH13629 to AAH13632  
XX represent oligonucleotides, all of which are used in the exemplification  
XX of the present invention.  
SQ Sequence 1045 AA;  
Query Match 26.6%; Score 3880.5; DB 22; Length 1045;  
Best Local Similarity 67.7%; Pred. NO. 9.3e-208;  
Matches 716; Conservative 165; Mismatches 155; Indels 21; Gaps 5;  
Qy 1686 MEKSTIDTIDENSRRDDVTITNLSSVEASVTYVQDVNSYQLCNGNGLRVYANGMGI 1745  
Db 1 mdkaitvdiesssreedsvitsnlssidsfycmvqdqrlrnsyqigdygsrlriysaglds 60  
Qy 1746 SFHSEPHVLAGTITPTIGRCNLSLPMWENLREKEQIKGVITFGKRLRVHGRNL 1805  
Db 61 hyqtpephvlagtanptvakrmtlpgengqlvverfkeaqgkvngvfgkrlvngnrl 120  
Qy 1806 LSIDYDRNIRTEKIYDDHKRFTLRIYDQVGRPFUWLPSGLAAVNSVFFPNRGLAQOR 1865  
Db 121 lsvdfdrttkttekyidhdkrflrlaydtsghptlwlpskklmavnvtyssgqiaslqr 180

QY 1866 GAMSERTDIDKQRIVRMPADGKWSYSLDKSMVLLLSQSQRYIFEYDSDRLAVTM 1925  
Db 181 gttsekvdvqggrivsvrvtadgktwaytylekmsvlllshsqryifeymwdrleaitm 240  
QY 1926 PSVARHSMSTHTSGYIRNLYNPPESNASVIFDYSDDGRILKTSFLGTGRQVFKYKGLS 1985  
Db 241 psvarhsmqctiralggyrnllynpesnaslftdyneeglllqltafgrsvrfkvrirt 300  
QY 1986 KLSEIYVDSTAVTFGDEYDTEVLKMWNLQSGFSGTIRYRKIGPLVDKQIYRFESEGMVN 2045  
Db 301 rlseilydstvstfdetavgiktnlqsdgfictiryqipldrqifrfesdgmvn 360  
QY 2046 ARDYYTHDSFRIASIKPVISETPLVDLYRYDEISGKVEHFGKGVYIYDINQIITTA 2105  
Db 361 arfdysy-dnsfrvmqgvinetpidlyfdidsgkveqfgvgfyvdydnqlaist 419  
QY 2106 VMTLSKHFDTHGRKEVOYEMFSLMVMYVQYDSMGRVTKRELKLGCPYANTTKYTYDYD 2165  
Db 420 vmytkhfdahgrikeiqeifrslymwticidydmgrvctkreikgpfanttkyayeyd 479  
QY 2166 GDGQLQSVAVNDRTWRYSDYDLNGLNLLNPGNSVRLMPLRYDLRITRLGDVQYKIDD 2225  
Db 480 vdsglqvtvlineklmwrnydnglnhlnpsnsarilplydrldritrlgdvqyrlde 539  
QY 2226 DGYLCQRGSDIFEPYNSKGLTRAYNKSAGWSVQVRYDGVGRRASYKTNLGHILQYFYSD 2285  
Db 540 dgflrqtgeifeysskglltrvysksgwtviyrydglgrvssktslghlqfyadl 599  
QY 2286 HNPTRITHVYHNSSETLSYLDLQHLFAMESSGREYVVASDNTGCTPLAVESINGLMI 2345  
Db 600 tpyrlichvynhasseltslydqlghlfameissgdefyiasdntgctplavfssnglmi 659  
QY 2346 KQLQYTAAYGEIYDSDNPDQFQWGFHGLYDPLTKLVHFTQRYDYVLGAWTSPDYTMWK 2405  
Db 660 kqlyctaygeiyfedsnidfvlqfghglvdpklklhfgedydlag-rtwtpdieiwk 719  
QY 2406 NVGKEPAPPNLYWFKSNPNLSSELDLKNYVTDVKSMLVMFGFQLSNLIPIGFPRAKMYFVP 2465  
Db 720 rikdpapfnlymfrnnpaskihdvkdytdvnswlvtfgfhlhnaipgfpvppkfdte 779  
QY 2466 PPELSESOASENGOLITGVQOQTERHNOAFMAL-EGQVITTKKLHASIREKAGH- 2518  
Db 780 pseyelksqgwdlplpffvqqvraqakafislgmaevqv-----srrraggaqs 831  
QY 2519 --WFAPTPTIIGKIMPAIKGRTVTVGVSSIASDSRKVASVLNNAYLLDKMHYSIEGKD 2576  
Db 832 wlfatvksligkvmlavsggrvqtnvlnianedcikaavlnnafylenlhftiegd 891  
QY 2577 THYFVKIGSADGLVLTGTTIGRKVLESGVNVVVSQPTLLVNGRTRFRFTNIEFOYSTLLL 2636  
Db 892 thyfiktttpesdltlrltsgrkalenglnvtvsgsttvngtrrfadvemqfgalal 951  
QY 2637 SIRYGLTPDPLDEKARVLQDQARALGTAWAKEQKARDGREGSLRWTEGEKQOLLSTG 2696  
Db 952 hvrygm---tldekarileqarqlarawareqqrdrdgegarlwtgeqkrqlisag 1008  
QY 2697 RVQGYEGYVLPVEQYPELADSSSNIQFLRQNMGR 2733  
Db 1009 kvagdygyvlsveqypeladsannqlrqsgeigr 1045  
RESULT 6  
AAB93294  
ID AAB93294 standard; Protein; 964 AA.  
XX  
AC AAB93294;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human protein sequence SEQ ID NO:12355.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX

OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
PF  
XX 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI; 2001-318749/34.  
XX  
CC Primer sets for synthesizing polynucleotides, particularly the 5602  
full-length cDNAs defined in the specification, and for the detection  
and/or diagnosis of the abnormality of the proteins encoded by the  
full-length cDNAs -  
PS Claim 8; SEQ ID 12355; 2537pp + CD ROM; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602  
full-length cDNAs defined in the specification. Where a primer set  
comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
to the complementary strand of a polynucleotide which comprises one of  
the 5602 nucleotide sequences defined in the specification, where the  
oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
of an oligonucleotide comprising a sequence complementary to the  
complementary strand of a polynucleotide which comprises a 5'-end  
sequence and an oligonucleotide comprising a sequence complementary to a  
polynucleotide which comprises a 3'-end sequence, where the  
oligonucleotide comprises at least 15 nucleotides and the combination of  
the 5'-end sequence/3'-end sequence is selected from those defined in  
the specification. The primer sets can be used in antisense therapy and  
in gene therapy. The primers are useful for synthesizing polynucleotides,  
particularly full-length cDNAs. The primers are also useful for the  
detection and/or diagnosis of the abnormality of the proteins encoded by  
the full-length cDNAs. The primers allow obtaining of the full-length  
cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
represent oligonucleotides, all of which are used in the exemplification  
of the present invention.  
XX  
SQ Sequence 964 AA;  
Query Match 24.9%; Score 3628.5; DB 22; Length 964;  
Best Local Similarity 69.0%; Pred. No. 9.4e-194;  
Matches 673; Conservative 139; Mismatches 143; Indels 21; Gaps 5;  
QY 1767 ISLPMENGLNSIEWLRKEQIKGVITFGKRLRVHGNLLSIDYDRNIRTEKIYDDHRKF 1826  
Db 1 mtlpgngqnlvewrfkdaqkvnvfgkrlrvngnllsvdfrtktekiyddhrkf 60  
QY 1827 TLRIYDQVGRPFLWLPLSSGAAVNVSYFFNGRLAGLQAGMSERTDIDKQGRIVSRMFA 1886  
Db 61 llriaydtsghptlwpssklmavnvtystgqiaslqrgttsekvdvqggrivsvrfa 120  
QY 1887 DGKWSYSYLDKSMVLLLSQSQRYIFEYDSDRLAVTMPSVARHSMSTHTSGYIRNLY 1946  
Db 121 dgkwsytylksmvlhllhsqryifeymwdrleaitmptsvrhtmtqirsigyryni 180  
QY 1947 NPPEASNASVIFDYSDDGRILKTSFLGTGRQVFKYKGLSKLSEIYVDSTAVTFGDEYD 2006  
Db 181 nppeasnasiitdyneeglllqltafgrsvrfkvrirtslrslldstsvrtydetag 240

QY	2007	VLKMYNLQSGFSCTIRYRKIKGLVDKQIYRSEBGMVNAERDYYTHDNSFRIASIKPVI	2066
Db	241	vlktnvlnqsgdfictiryqglpdlrqlfrfseggmgnarfdisy--dnsvrvtsmggvi	299
QY	2067	SETPLVDLIRYDEISGKVEHFGKGVIIYDINQIITAVMTLSKHFDTHGRKEVQVEM	2126
Db	300	netplidlyqfddlsqgvegfkgfgyvdinginqiaistavmtytkhfdahgrikeigvei	359
QY	2127	FRSLMYMTVOYDSMRVKKRLKGLPVANTTKYNYDVGQLOLSVAVNDRPTWRYSD	2186
Db	360	fslmywtlgydmgrvckreiklpgfantckeyeydvdgqlqtcvynlnekimwrynyd	419
QY	2187	LGNLHLLNPGNSVRLMLPLRDLDRITRLGLDVOYKIDDDGYLCQRGSDIFEYNSKGLLT	2246
Db	420	lgnlhllnpnsarlplyldrlrcltrlgdvqyrldegflrgteifeysskgilt	479
QY	2247	RAYKASGWSVOYRYDVGVRASRYTNLGHLOYPFSLHNPTRTHVYNINSNSETLSLY	2306
Db	480	rvysksgwtviyrydgygrrvssktslqhlqfyadltyptrithvynhsseitsly	539
QY	2307	YLOQLHLFAMESSEGEYVVASDNTGTPLAVFSINGLMIKOLQYATYGEIYYDSNPDQOM	2366
Db	540	ydlqqlhlfamelssgedefiasdntcgtplavfssnglmkqiqtaygeiyfidsnidfql	599
QY	2367	VIGFHGGYDPLTKLVLHFTQRDYDVLAGRWTSPDYTMKNVKGEPAPFNLYMFKSNPLS	2426
Db	600	vlgfhgglydpltklhfgerdydlagrwtpdiekwkrgkdpapfnlymfrnnpas	659
QY	2427	SELDLKNVYDVKSLVWFGQLSNIIPGFPAKMYFVPPPELSESOASENQSLITGVQ	2486
Db	660	kldvkdyltdvnswlvtfgfhnalppfpvkdltlepsyvelvksqgdldpifgvq	719
QY	2487	QTTERRHNOAFMAL----EQGVITTKKLHASIREKAGH----WFATTTPIIGKIMFAIKE	2537
Db	720	qqvraqakafislgmaevq-----srraagagswlfatvkslkgvmlavsq	771
QY	2538	GRVTTGVSSIASERDRKVASVLNNAYYLDKMHYSTEGKDTHTFYKIGSADGLVTLGTTI	2597
Db	772	grvqtnvlnianedekvaavlnnafylenhftiegdktthyfiktttpesdlglrlts	831
QY	2598	GRKVLSEGVNVTVSQPTLLVNGTRRFTNIEFOYSTLLSIRYGLTPDTLDEEKARVLDO	2657
Db	832	grkalenglnvtvsgstvtvngtrfrfdavemqfgalalhvyrgm---tldceekarileq	888
QY	2658	ARQALGTAWAKEQKARDREGSRSLWTEGEKQQLLSTGRVQYEGYVLPVEQYPELAD	2717
Db	889	arqalarawareqrvrdgeegarlwtegekrqllsagkvgydygyvylsveqypelad	948
QY	2718	SSSNTQFLRQNEGKR	2733
Db	949	sannlqflrqseigr	964
RESULT	7		
AAW79679			
ID	AAW79679	standard; Protein; 1015 AA.	
XX	AC	AAW79679;	
XX	AC	AAW79679;	
DT	06-NOV-2001	(first entry)	
DE	Human	protein SEQ ID NO 3325.	
KW	Human;	cytokine; cell proliferation; cell differentiation; gene therapy;	
KW	vaccine;	peptide therapy; stem cell growth factor; haematopoiesis;	
KW	tissue growth factor;	immunomodulatory; cancer; leukaemia;	
KW	nervous system disorder;	arthritis; inflammation.	
OS	Homo sapiens.		
XX	PN	W0200157190-A2.	
XX	XX		

PD	09-AUG-2001.		
XX	05-FEB-2001; 2001WO-US04098.		
XX	03-FEB-2000; 2000US-0496914.		
PR	27-APR-2000; 2000US-0560875.		
PR	20-JUN-2000; 2000US-0598075.		
PR	19-JUL-2000; 2000US-0620325.		
PR	01-SEP-2000; 2000US-0654936.		
PR	15-SEP-2000; 2000US-0663561.		
PR	20-OCT-2000; 2000US-0693325.		
PR	30-NOV-2000; 2000US-0728422.		
XX	(HYSE-) HYSEQ INC.		
PA	Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;		
PI	Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;		
PI	Xue AJ, Yang Y, Wejhrman T, Goodrich R;		
XX	WPI; 2001-476283/51.		
DR	N-PSDB; AAK52812.		
XX	Nucleic acids encoding polypeptides with cytokine-like activities,		
PT	useful in diagnosis and gene therapy -		
XX	Claim 20; Page 307; 6221pp; English.		
PS	The invention relates to polynucleotides (AAK51456-AAK53435) and the		
CC	encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to		
CC	cytokine, cell proliferation or cell differentiation or which may induce		
CC	production of other cytokines in other cell populations. The		
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or		
CC	peptide therapy. The polypeptides have various cytokine-like activities,		
CC	e.g. stem cell growth factor activity, haematopoiesis regulating		
CC	activity, tissue growth factor activity, immunomodulatory activity and		
CC	activation/inhibin activity and may be useful in the diagnosis and/or		
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and		
CC	inflammation.		
CC	Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666		
CC	(AAM80020) are omitted as the relevant pages from the sequence listing		
CC	were missing at the time of publication.		
XX	Sequence 1015 AA;		
SQ	Query Match 19.5%; Score 2837; DB 22; Length 1015;		
	Best Local Similarity 53.6%; Pred. NO. 1.4e-149;		
	Matches 542; Conservative 191; Mismatches 267; Indels 12; Gaps 5;		
QY	1707	TNLSVEASYTVVQ--DQVRNSYQLCNGNGLTRVWYANGGISFHSPEHVLAGTITFTIGRC	1765
Db	9	sslpssssfallstentgrtyrvnpgdglrvtfasgmeiglssephilagavnpnlkgc	68
QY	1766	NISLPMENGLNSIEWRLRKEQIKGKVTIFGKRRLRVHGNLLSIDYDRNIRTEKIYDDHRK	1825
Db	69	nlsipgehnanlisvxxgeqgcaxnvfhlsfsxhnnnlisidfdhtrfgkydyddhrk	128
QY	1826	FTLRIYDQVGRPEFLWLPSSGLAAVNVSYFFNGRLAGLQRGAMSBERTIDKQGRIVSRMF	1885
Db	129	ftlriylqdtgrplwspvsrynevnityspglvtfigrqtwnmekmeydqsfksplxl	188
QY	1886	ADGKWSYSYLDKSMVLLQSQORQYIFEYDSSDRLLAVTMPSPVARHSMSTHTSGYIRNI	1945
Db	189	lsilcysafsvfsqsmlllshsrryifeydqpdcillsvtmvsmvhsiqtlslsvgyyni	248
QY	1946	YNPPESNAVIFDSDGGRILKTSPLNGRQVYFYKYGKLSKLSEIVDYSTAVTFYDETT	2005
Db	249	ytppdsstsfldqyrdgrliqlthlgcrrvrykytkqarlseviydttdvtlyceess	308
QY	2006	GVLMKVNLSQSGFSCTIRYRKIKGLVDKQIYRSEBGMVNAERDYYTHDNSFRIASIKPVI	2065
Db	309	gdlssdstliaxlltvtvlpagpligrqifrfseglvnmardysy--nnfrvtmsmqav	366



QY 2578 HYFVKTSADGLVLTGTTIGRKVLSEGVNVTVSQPTLLVNGTRTRFTNIEFYSTLLS 2637  
Db 473 hyfikttesdgltrltslgskalenginvtsqsttvngtrrrfadvemqfgaalh 532  
QY 2638 IRYGLTPDILDEKARVLDOARQALGTAWAKEQKARGREGSLWTEGEGQOOLLSTGR 2697  
Db 533 vrygm---tldeekarileqarqralarawareqqrdrdeegarilwtegekrqlisagk 589  
QY 2698 VQGYEGYVLPVQVPELADSSNQFLRQNMGR 2733  
Db 590 vqgygyvylsvqypeladsannqlfrqselgrf 625  
RESULT 9  
AAU00392  
ID AAU00392 standard; Protein; 429 AA.  
XX  
AC AAU00392;  
XX  
DT 04-JUL-2001 (first entry)  
Human secreted protein, POLY4.  
KW Human secreted protein; therapeutic; diagnostic; human; cancer.  
OS Homo sapiens.  
XX  
PN WO200119856-A2.  
XX  
PD 22-MAR-2001.  
XX  
PF 13-SEP-2000; 2000WO-US25106.  
XX  
PR 13-SEP-1999; 99US-0153629.  
PR 16-SEP-1999; 99US-0154520.  
PR 20-SEP-1999; 99US-0154762.  
PR 13-OCT-1999; 99US-0159231.  
PR 12-SEP-2000; 2000US-0659634.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shinkets RA, Fernandes E, Herrmann JL, Liu X, Yang M, Boldog FL;  
XX  
DR WPI; 2001-244781/25.  
DR N-PSDB; AAS01213.  
XX  
PT New POLYX polypeptide useful for treating or preventing a POLYX  
XX associated disorder, e.g. cancer  
XX  
XX Claim 9; Page 11-13; 152pp; English.  
CC The sequence represents the amino acid sequence of human secreted  
CC protein, POLY4. POLYX nucleic acids, polypeptides and antibodies to POLYX  
CC can be used for treating or preventing a POLYX associated disorder in a  
CC subject, preferably a human. These can be used in the manufacture of a  
CC medicament for treating a syndrome associated with a human disease  
CC selected from a POLYX-associated disorder, where the therapeutic is a  
CC POLYX polypeptide, a POLYX nucleotide or a POLYX antibody. They may also  
CC be used to screen for a modulator of activity, or latency, or  
CC predisposition to a POLYX associated disorder, e.g. cancer.  
XX  
SQ Sequence 429 AA;  
Query Match 14.6%; Score 2122.5; DB 22; Length 429;  
Best Local Similarity 98.8%; Pred. No. 2.8e-110;  
Matches 398; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
QY 162 KSDDEGRPIPTSPSLPSAQLPSSHNPPVSCQMPILDSNTSHQIMDTPDEEFSN 221  
Db 10 kpsaeagrpiptspslpsaqlpsshnppvscqmpildntshqimdtndpdeefspn 69

QY 222 SYLLRACSGPQOASSGGPPNHHHSQSLRPLPPPHNHTLSHHSSANSNRSLNRRSQ 281  
Db 70 syllracsgpqqassgppnhsqslrplppphnhtlshhssanslnrsltnrrsq 129  
QY 282 IHAPAPAPNDLATTPEVOLQDSWVLSNVPLETRHFLFKTSSGSTPLFSSSSPGYPLTS 341  
Db 130 ihapapapndlattpesvqlqdswnlnsnvpletrhflfktsagstplfsssgpyplts 189  
QY 342 GTVYTPPPRLPRNTFSRKAFKLKPKSKYCSWKAALSIAAALLAILLAYFIVPWSLK 401  
Db 190 gtvytppprllprntfsrkafklkpkskycswkaalsaiaaallaillayfivpwslk 249  
QY 402 NSSIDSGEAEVGRRTVQVPPGVFWRSQIHSQPOFLKFNISLGKDALFGVYIRRLGPPS 461  
Db 250 nssidsgaeavgrvtqvppgvfwrshisqpqlknislgkdalfgvyirrlgpps 309  
QY 462 HAQYDFMERLDGKESVSVESPRERSIOTLVQNEAVFYQYLDVGLWHLAFYNDGKDKEM 521  
Db 310 haqydfmerldgkesvsvesprersiotlvqneavfyqyldvglwhlafyndgdkem 369  
QY 522 VSFNTVVL-DSVQDCPRNCHNGECVSGVCHCFPGFGLGADCAK 563  
Db 370 vsfntvvlldsvqdcprnchngecvsgvchcfpgflgadca 412  
RESULT 10  
ABG07028  
ID ABG07028 standard; Protein; 746 AA.  
XX  
AC ABG07028;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #7019.  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS71215.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity  
XX  
PS Claim 20; SEQ ID No 37387; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical



```
Query Match      13.98; Score 2022.5; DB 22; Length 746;  
Best Local Similarity 43.1%; Pred. No. 2.3e-104;  
Matches 466; Conservative 38; Mismatches 72; Indels 505; Gaps 17;
```

```
812 DNEGDGLVDCLDPCCLQSAQNLSLLCRGSRDPLDIIOGGOTDWPVAKSFYVDRIKLLAGK 871  
l : |||||  
Db dsvsdglvdcldpcclqsaqnlsllcrgrsrdpdlilgggtcdwpvaksfydrkllagk 129  
  
Qy DSTHIIIPGENPFSNLSVLIRGOVVTTDTGPLVGNNVSFVKYPKYGYTTRQDGTFDLIA 931  
|||||  
Db dsthiiipgenpfns-----ehls 148  
  
Qy NGGASLILHFPRAPFMPSOERTVWLPWNFSFYAMDTLVMKTENSIPOCDLSCGFVRDPPII 991  
||| | :  
Db pghashlw-----rsisrhaw-----hgndlisidydnirtekiyd 187  
  
Qy SSPLSTFFSAAQPONPIVPETQVLHEEIEELPGSNVKRLYSRTAGYKSLLKITWTQTSTV 1051  
+ : : : : : rilydqvrp----- 204  
Db dhrrftl-----rilydqvrp----- 204  
  
Qy PLNIIRVHLMAVEGHFLFKSQASPNLASTFINDKTDAYGORVYGSLDAVVSVGFYEYET 1111  
| : | : || | : :  
Db 205 -----flwlpss-----gl-aavnvysffn- 223  
  
Qy CPSLIIMEKRATLLOQFELDPSNLGGWSLDKHHLNVKSGILHKCT-GENOFLTQQPAII 1170  
| | | : | : | :  
Db 224 -----griag-----lrgamserdtidkggriv 247  
  
Qy TSMGNRRRSISCPSCNGLAEGNKLLAPVALAVGIDGSLVXVDFNYRIRPPSPNRVTISI 1230  
: : ! : | : | : :  
Db srmfadgk-----vwysyl----- 262  
  
1231 LELRNKEPKHSNPAPHKKYYLAVDPVSGSXVSDTNRRRIYRVKSLSGTKDLAGNSEVWAG 1290  
263 ----- 262  
  
Qy TGEQLPFDEARCGDGGKAIDATLMSPRGIAVDKNGLMYFDVATMIRKVQDNGIISTLLG 1350  
| : |  
Db 263 -----dkdgcg----- 269  
  
Qy SNDLTAVRPLSCDSMDVAQVRLEMPDTLAVNPMDNSLYLVLENNVILRITENHQVSIAG 1410  
270 ----- 269  
  
Qy RPMHCQVPGIDYSLKLAHSALESASAIAISHTGVLYITEFDEKKINLRQVTTNGEIC 1470  
||| | : ||  
Db 270 -----kstgcrl-----rkscr----- 280  
  
Qy LLGAASDCCKDNVCNCYSDDDAYATDALNSPSSLAVAPDGTIYTADLGNIIRAVS 1530  
||  
Db 281 -----cyqrqk-----sdsgni----- 292  
  
Qy KNKPVLNAFNQYEAAASPEQEELYVENADGIHOYTIVSLTVGETVLYNFYTSNDMTVELIDN 1590  
|| |||||  
Db 293 -----afiqyeaaaspeeqlyvnaadgihoaytvtgetvlylfystdndvelidn 345  
  
Qy NGSULKIRRSSGMPRHLLMPDNQIITIUTVGTNGGLKVVYSTQNLEGMLTYDQGNTGLLAT 1650  
|||||
```

DR N-PSDB; AAS14084.  
XX Novel polypeptides designated as FCTR polypeptides, useful in  
PT detection, prevention and treatment of a broad range of pathological  
PT states  
XX  
PS Claim 1; Page 33; 215pp: English.  
XX  
XX The invention relates to human FCTR polypeptides, FCTR1-FCTR7, and the  
CC nucleic acids encoding them. These sequences are useful for the treatment  
CC or prevention of numerous disorders including myelogenous leukemia,  
CC carcinomas, melanomas, gliomas, astrocytomas, congenital neonatal  
CC autoimmune thrombocytopenia, neurological disorders, neurodegenerative  
CC disorders, nerve trauma, familial myelodysplastic syndrome,  
CC Charcot-Marie-Tooth neuropathy, demyelinating Gardner syndrome, familial  
CC myelodysplastic syndrome, mental health conditions, immunological  
CC disorders, allergy and infection, bronchial asthma, Avellino type  
CC eosinophilia, lung diseases, reproductive disorders, infertility, male  
CC and female reproductive disorders, deafness, glycoprotein Ia deficiency,  
CC desmoid disease, turtot syndrome, liver cirrhosis, hepatitis C, gastric  
CC disorders, pancreatic diseases such as diabetes, Schistosoma mansoni  
CC infection Spino cerebellar ataxia, Plasmodium falciparum parasitaemia,  
CC Corneal dystrophy-Greenow type I, Corneal dystrophy-lattice type I and  
CC Reis-Bucklers corneal dystrophy. This sequence represents FCTR3a, a  
XX neurestin-like protein.  
XX  
SQ Sequence 381 AA;  
  
Query Match 13.2%; Score 1931; DB 22; Length 381;  
Best Local Similarity 98.9%; Pred. No. 1.le-99;  
Matches 365; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
Qy 162 KSDDENGRPIPTTSSPLLPSAQLPSSHNPPVSCOMPLDLSNTSHQIMDTNPDEEFSN 221  
Db 10 kpsaeagrpipttsspsllpsaqilpsshnpvpsqmpdlldntshqimdnpeefspn 69  
  
Qy 222 SYLLRACSGPOQASSSGPPNHHSSQSTLRPPPLPPPHNHTLSHHSSANSLSNLSNRRSQ 281  
Db 70 syllracsgpqgassgppnhhsqstlrpppphnhhtlshhssanslnrnslnrrsq 129  
  
Qy 282 IHAPAPANDLATTPEVSQLODSWVLNSVPLETRHFLEFKTSSGSTPLFSSSSCYPLTS 341  
Db 130 ihapapandlattpesvqlqdswwlnsnvpletrhfiktsqstplfsssspgypilts 189  
  
Qy 342 GTVYTPPRLPLRNTFSRKAFKLLKPKSKYCSWKCAALSAIAAALLAILLAYFIVPWSLK 401  
Db 190 gtvytpprllprntfsrkafkllkpkskycskwcaalsaiaaallaillayfivpwslk 249  
  
Qy 402 NSSIDSGAEVGRRTQEVPPGVFWRSQIHISQPOFLKFNISLCKDALFGVYIRGLPPS 461  
250 nssidsgaeagrvtqevppgvfwrsgqihisqpfkfnislgkdalfgvyirglpps 309  
  
Qy 462 HAQYDFMERLDGKWKSVESPRRRSIQTLVQNEAVFVQYLDVGLWHLAFVNDGKDKEM 521  
Db 310 haqydfmerldgkewsvesprrrsiqtlvqneavfvqyldvglwhlafyndgdkkem 369  
  
Qy 522 VSFNTVVLID 530  
Db 370 vsfntvvlid 378  
  
RESULT 12  
AAB61141  
ID AAB61141 standard; Protein; 381 AA.  
XX  
AC AAB61141;  
XX  
DT 30-MAR-2001 (first entry)  
XX  
DE Human NOV11 protein.  
XX  
KW Human; NOVX; antinflammatory; cytostatic; neuroprotective;

KW cerebroprotective; immunomodulator; vulnerary; vasotropic; gene therapy;  
KW hyperplasia; tumour; restenosis; psoriasis; Dupuytren's contracture;  
KW diabetes; rheumatoid arthritis; cerebral oedema; Alzheimer's disease.  
XX Homo sapiens.  
XX  
PN WO200075321-A2.  
XX  
PD 14-DEC-2000.  
XX  
XX 01-JUN-2000; 2000WO-US15303.  
XX  
PR 03-JUN-1999; 99US-0137322.  
PR 16-MAR-2000; 2000US-0189810.  
PR 22-MAR-2000; 2000US-0191158.  
PR 30-MAR-2000; 2000US-0193086.  
PR 31-MAY-2000; 2000US-0137322.  
XX (CURA-) CURAGEN CORP.  
XX  
PA Shmkets RA, Fernandes E, Herrman J, Vernet C;  
PI  
XX  
XX WPI: 2001-102403/11.  
DR N-PSDB; AAF27859.  
DR  
XX  
PT New NOVX polypeptides and polynucleotides, useful in gene therapy, as a  
PT diagnostic marker, protein therapeutic, antibody or small molecule drug  
PT target for treating immune, proliferative and metabolic diseases and  
PT wound healing  
XX  
XX Claim 1; Page 42-44; 194pp: English.  
PS  
XX  
XX The present sequence is a new isolated polypeptide (NOVX). The NOVX  
CC polypeptides, NOVX nucleic acids, and anti-NOVX antibodies are useful for  
CC treating or preventing NOVX-associated disorders. They are also useful  
CC for determining the presence of or a predisposition to a disease  
CC associated with altered levels of the NOVX polypeptide or nucleic acid.  
CC These NOVX-associated disorders include hyperplasias, tumours,  
CC restenosis, psoriasis, Dupuytren's contracture, diabetic complications,  
CC rheumatoid arthritis, cerebral lesions, diabetic neuropathies, cerebral  
CC oedema, senile dementia or Alzheimer's disease. The NOVX polynucleotides  
CC are especially useful in gene therapy. Specifically, NOVX is useful as  
CC a diagnostic marker or prognostic marker, protein therapeutic and  
CC antibody target or small molecule drug target to treat disorders in the  
CC immune response pathway, thyroid and metabolic diseases, bone metabolic  
CC disorders, diseases of the pancreas (e.g. diabetes or digestive  
CC disorders), proliferative diseases, or tissue regeneration and  
CC development (e.g. wound healing or treatment of burns).  
XX  
SQ Sequence 381 AA;  
  
Query Match 13.2%; Score 1931; DB 22; Length 381;  
Best Local Similarity 98.9%; Pred. No. 1.le-99;  
Matches 365; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
Qy 162 KSDDENGRPIPTTSSPLLPSAQLPSSHNPPVSCOMPLDLSNTSHQIMDTNPDEEFSN 221  
Db 10 kpsaeagrpipttsspsllpsaqilpsshnpvpsqmpdlldntshqimdnpeefspn 69  
  
Qy 222 SYLLRACSGPOQASSSGPPNHHSSQSTLRPPPLPPPHNHTLSHHSSANSLSNLSNRRSQ 281  
Db 70 syllracsgpqgassgppnhhsqstlrpppphnhhtlshhssanslnrnslnrrsq 129  
  
Qy 282 IHAPAPANDLATTPEVSQLODSWVLNSVPLETRHFLEFKTSSGSTPLFSSSSCYPLTS 341  
Db 130 ihapapandlattpesvqlqdswwlnsnvpletrhfiktsqstplfsssspgypilts 189  
  
Qy 342 GTVYTPPRLPLRNTFSRKAFKLLKPKSKYCSWKCAALSAIAAALLAILLAYFIVPWSLK 401  
Db 190 gtvytpprllprntfsrkafkllkpkskycskwcaalsaiaaallaillayfivpwslk 249  
  
Qy 402 NSSIDSGAEVGRRTQEVPPGVFWRSQIHISQPOFLKFNISLCKDALFGVYIRGLPPS 461

Db 250 nssidsgeaevrrvtqevppgvtrsqihisqpfkfnslgkdaifgvyirgpps 309  
Qy 462 HAQYDFMERLDGKESWVESPRRRSITQTLVQNEAVFOYLDVGLWHLAFYNDGKDEM 521  
Db 310 haqydfmerldgkeswvesprrrsigtltlvqneavfgyldvglwhlafyndgkdem 369  
Qy 522 VSFNTVVLD 530  
Db 370 vsfntvvld 378  
  
RESULT 13  
ABG04674  
ID ABG04674 standard; Protein; 402 AA.  
XX AC ABG04674;  
  
13-FEB-2002 (first entry)  
Novel human diagnostic protein #4665.  
Human; chromosome mapping; gene mapping; gene therapy; forensic;  
food supplement; medical imaging; diagnostic; genetic disorder.  
OS Homo sapiens.  
PN WO200175067-A2.  
PD 11-OCT-2001.  
PF 30-MAR-2001; 2001WO-US08631.  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
PI Drmanac RT, Liu C, Tang YT;  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS68861.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
Claim 20; SEQ ID NO 35033; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 402 AA;  
SQ

Query Match 11.7%; Score 1708.5; DB 22; Length 402;  
Best Local Similarity 92.7%; Pred No 3.1e-87;  
Matches 331; Conservative 4; Mismatches 9; Indels 13; Gaps 2;  
  
Qy 1899 SMVLLQSQRQYIFEXDSSDLLAVTMPSVARHSMSTHTSIGYIRNIYNPPESNASVFD 1958  
Db 27 smvllqsqrqyifeydssdrllavtmppsvarhsmsthtsigyirniynppesnasvfd 86  
  
Qy 1959 YSDGGRILKTSFLGTGRQVFKYKGLSKLSEIYVYDSTAVTFGVDYETTGVLKMWNLQSGGF 2018  
Db 87 ysdggrilksfllgtgrqvfykglsklseiyydstavtfgydettgvlkmvnlqsggf 146  
  
Qy 2019 SCTIRYRKIGPLVDKQIYRFSEEGMVARFDYTVHDSFRIASIKPVISETPLPVDLYRY 2078  
Db 147 sctiryrkigplvdqkiyrfseegmvarfodytyhdnsfriasikpvisetpplvdyry 206  
  
Qy 2079 DEISGKVEHFGKFGVIYIDINQIITAVMTLSKHFDHGRIVEY-QYEMFRLMYWMTVQ 2137  
Db 207 deisgkvehf-----fittaeamtlskhfdhgriveyemfrrlmywmtvq 254  
  
Qy 2138 YDSMGRYIKBELKGLGPYANTTKYDYDGGQLOQSVAVNDPRTWRYSDYDLNGLHLLNPG 2197  
Db 255 ydsmgrvirkelkigpyanttktydydggdqloqsvacndrptwryseyelngnlhllnpg 314  
  
Qy 2198 NSVRLMPLRYDLRDRITRLGQVQYKIDDDGYLCQSGSDIFEYNSKGLLTRYNNKASG 2254  
Db 315 nsvrlmplrydlrdritrlgqvgqykiddgylcqsgsdifeynfgaptktsqsg 371  
  
RESULT 14  
ABG04673  
ID ABG04673 standard; Protein; 242 AA.  
XX AC ABG04673;  
XX 13-FEB-2002 (first entry)  
XX Novel human diagnostic protein #4664.  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
XX Homo sapiens.  
XX WO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US08631.  
XX 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
XX N-PSDB; AAS68860.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX Claim 20; SEQ ID NO 35032; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 402 AA;  
SQ



Db 627 kgka-aaacd 637  
:| : ||||

Search completed: September 18, 2002, 11:08:02  
Job time: 307 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 11:05:30 ; Search time 48.75 Seconds  
(without alignments)  
5386.916 Million cell updates/sec

Title: US-09-800-198-8  
Perfect score: 14581  
Sequence: 1 MDVKDRRHSRLTRGRCGEC.....ELADSSNIQFLRQNMGR 2733  
Scoring table: BLASUM62  
Gapop 10.0 , Gapext 0.5

rched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9522	65.3	2825	2 T14271	Doc4 protein, stre
2	4465	30.6	849	2 T46253	hypothetical prote
3	3916.5	26.9	2515	2 S47008	tenascin-like prote
4	3637.5	24.9	2406	2 A54148	odz protein - frul
5	2208.5	15.1	2531	2 T16743	hypothetical prote
6	1171	8.0	782	2 A61625	tenascin-like prote
7	726	5.0	1810	1 A32230	tenascin precursor
8	680	4.7	2201	2 A32160	tenascin-C - human
9	679.5	4.7	2019	1 JQ1322	tenascin precursor
10	656	4.5	1746	1 S19694	tenascin precursor
11	632.5	4.3	4006	2 T09070	probable tenascin
12	627.5	4.3	184	2 T12457	hypothetical prote
13	616	4.2	3566	1 A40701	tenascin-X precurs
14	604	4.1	647	2 A43902	tenascin - eastern
15	590	4.0	4135	2 T42629	tenascin-x - bovin
16	463.5	3.2	2531	2 A46019	Notch-1 protein
17	462.5	3.2	2531	2 S18188	notch protein homo
18	452.5	3.1	3191	2 T22945	hypothetical prote
19	450.5	3.1	2318	2 S45306	notch 3 protein -
20	439	3.0	2555	2 A40043	notch protein homo
21	438.5	3.0	2321	2 S78549	notch3 protein - h
22	437.5	3.0	1064	2 A40136	fibropellin Ia - s
23	437	3.0	1620	2 T27283	hypothetical prote
24	434.5	3.0	1220	2 A56136	jagged protein pre
25	429.5	2.9	2471	2 A49128	cell-fate determin
26	428.5	2.9	2524	2 A35844	Xotch protein - Af
27	421.5	2.9	1203	2 A49175	Notch B protein -
28	421	2.9	1356	2 A45445	Janusin precursor
29	419.5	2.9	1574	2 T13954	MEGF6 protein - ra

30	417	2.9	2139	2 A35672	crumbs protein - f
31	417	2.9	2437	2 S42612	transmembrane prot
32	413	2.8	2703	1 A24420	notch protein - fr
33	412.5	2.8	1353	1 JH0675	restictin precurs
34	412.5	2.8	2531	2 T31070	notch homolog - se
35	405.5	2.8	2352	2 T30201	notch homolog prot
36	399	2.7	1295	2 A32901	gip1 protein precu
37	397	2.7	832	2 A31246	neurogenic protein
38	397	2.7	880	2 S00670	neurogenic repetit
39	395.5	2.7	1408	2 S16148	gene serrate prote
40	395	2.7	833	2 S19087	gene Delta protein
41	395	2.7	1111	2 T26972	hypothetical prote
42	391	2.7	1964	2 T09059	notch4 - mouse
43	389.5	2.7	1722	2 E89753	protein F11C7.4 [1
44	372.5	2.6	861	2 A48825	Notch homolog Motc
45	368.5	2.5	473	2 A56175	adhesive plaque pr

ALIGNMENTS

Doc4 protein, stress-induced - mouse  
N:Alternate names: odz protein homolog  
C:Species: Mus musculus (house mouse)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T14271  
R:Wang, X.Z.; Kuroda, M.; Sok, J.; Batchvarova, N.; Kimmel, R.; Chung, P.; Zinszner, EMBO J. 17, 3619-3630, 1998  
A:Title: Identification of novel stress-induced genes downstream of chop.  
A:Reference number: Z17951; MUID:98315054  
A:Accession: T14271  
A:Status: preliminary; translated from GR/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-2825 <MAN>  
A:Cross-references: EMBL:AF059485; NID:g3170614; PID:g3170615; PIDN:AAC31807.1  
C:Genetics:  
A:Gene: Doc4

Query Match 65.3%; Score 9522; DB 2; Length 2825;  
Best Local Similarity 61.6%; Pred. No. 0;  
Matches 1763; Conservative 421; Mismatches 511; Indels 168; Gaps 28;

Qy	1	MDVKDRR-HRSLTRGRCGECRYTSSLDSDCRVPTQKSYSSSETLKAYDHDSDRMHYGN	59
Db	1	MDVKRRPYKSLTRRR-DAERRYTSSSADSEEGKP-QKSYSSSETLKAYDQDARLAYGS	58
Qy	60	RVTDLIHRESDEFPRQGTNETLAEIGICEPS-PHRSGCYSDMGLHQYSLSTGSDADSD	118
Db	59	RVKDMVPQEAEEFCRTGNTFLRELGLGEMTPPHGTLRYTDIGLPHCGYSGAGSSADADLE	118
Qy	119	TEGMSPEHAIRLWGRGKSRSSGLSRNSALTLDSDNENKSDDENGRIPTSSPS	178
Db	119	ADTVLSPEHPVRLWGRSTRSGRSCLSRANSNLTDTEHEN---TETGAPL-HCSSAS	174
Qy	179	LLPSAQPLSSHNPPVS-CQPLLDNSTSHQIMDTPDEEPSNYSLLRACSGPQQASSS	237
Db	175	STPIEQSP	234
Qy	238	GPP-----NHHSQSRLRPLPPPHNHTLSHHH-SSANSINRNLNRRSQIHAPAPND-	291
Db	235	DHPSSLQNHPLRLP-PPPLPHAPHTNQHHAASINLRNNGFTPRSN-----PSPAPTDH	288
Qy	292	-----LATTPESVQLQDSVNLNSNPLETR-----	316
Db	289	SLSGEPPAGSAQEPHQAODNVLNSKIPVETRNLGKQPLCTQMDNLIEMDIFSASRRDG	348
Qy	317	-----HFLFKTSSGSTFLFSSSSPGYPLTSGTYVTPPRLPLPNTFSRKAFLKPKSKYC	371
Db	349	AYSDGHEFFK-PGCTSPFLCTTSPGYPLTSSVYSPPPPLPPLPRTFSRPAFLNKKPKSKYC	407





Db 2550 TDAMEPSYELVHTQMTQWMDNSKILGVOCEVQKQKAFVTLERFDQLYGSTIISCOQA 2609  
Qy 2511 SIREKAGHWFATTTPIICKGIMFAIKEGRVTTGVSSIASEDSRKVASVLLNAYYLDKWHY 2570  
Db 2610 PETKK-----PASSGIFGKGVKALKDRVTTDIISVANEDGRRRAALINNAHYLENLHF 2665  
Qy 2571 SIEGDKTHYFKVIGSADGLVLTGTTIGRKVLESVNVTVSQPTLLVNGTRRRRTNIEFQ 2630  
Db 2666 TIDGVDTHYFKVPGPSEGDLAILGLSGGRRRTLENGVNVTVSQINTMLSGRRRTYDIQLQ 2725  
Qy 2631 YSTLLSTRYGLTPDTLDEEKARVLDQARQALGTAWAKEQKARDGREGSLRTEGKQ 2690  
Db 2726 YRALCLNRYG---TVDDEKVRVLELQARAVRAQREOQLREGEGLRAWTDGEKQ 2782  
Qy 2691 OLLSTGRVQVGEYVYVPEQYPELADSSNIQFLRONEMGR 2733  
Db 2783 QVLTNTRVQVGDGFFVTSVEQYPELSDSANNIHFNRQSEMGRR 2825

RESULT 2  
T46253  
hypothetical protein DKFzp761F171.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C:Accession: T46253  
R:Ottewaeldeier, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, January 2000  
A:Reference number: 223031  
A:Accession: T46253  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-849 <AAA>  
A:Cross-references: EMBL:AL137500  
A:Experimental source: adult amygdala; clone DKFzp761F171  
C:Genetics:  
A:Note: DKFzp761F171.1

Query Match 30.6%; Score 4465; DB 2; Length 849;  
Best Local Similarity 100.0%; Pred. No. 7.4e-230;  
Matches 849; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1885 FADGKWSYSLDKSWLLLASQROYIFEYDSSDRLLAVTMPVARHSMSTHTSGYTRN 1944  
Db 1 FADGKWSYSLDKSWLLLASQROYIFEYDSSDRLLAVTMPVARHSMSTHTSGYTRN 60

Qy 1945 IYNPPESNASVIFDYSDDGRILKTSFLGTGRQVFKYKLSLSEIVYDSTAVTFGYDET 2004  
Db 61 IYNPPESNASVIFDYSDDGRILKTSFLGTGRQVFKYKLSLSEIVYDSTAVTFGYDET 120

Qy 2005 TGVLRVNLQSGGFSCTIRYRKIGPLVDKQIYRFSEEGMVNARFDYTHDNSFRASIATK 2064  
Db 121 TGVLRVNLQSGGFSCTIRYRKIGPLVDKQIYRFSEEGMVNARFDYTHDNSFRASIATK 180

Qy 2065 VISETPLVDLYRYDEISGVEHFGKFGYIYDINQIITAVMTLSKHPDTHGRKEVOY 2124  
Db 181 VISETPLVDLYRYDEISGVEHFGKFGYIYDINQIITAVMTLSKHPDTHGRKEVOY 240

Qy 2125 EMFRSLMVMVTQYDSMGVRITKRELKGLPYANTTKYTYDYGDLQLOSVAVNDRPTWYS 2184  
Db 241 EMFRSLMVMVTQYDSMGVRITKRELKGLPYANTTKYTYDYGDLQLOSVAVNDRPTWYS 300

Qy 2185 YDLNGLHLLNPGNSVRLMPLRYDLRDRITRLGDYQYKIDDDGYLCQSGOIFEYNSKGL 2244  
Db 301 YDLNGLHLLNPGNSVRLMPLRYDLRDRITRLGDYQYKIDDDGYLCQSGOIFEYNSKGL 360

Qy 2245 LTRAYNKASGWSQYRYDGVGRASRYKTNLGHLLQYFYSDLNPNTRITHVYHNSSEITS 2304  
Db 361 LTRAYNKASGWSQYRYDGVGRASRYKTNLGHLLQYFYSDLNPNTRITHVYHNSSEITS 420

Qy 2305 LYYDLOGLHFLAMESSGEYVYASDNTGTPLAVFSINGLMIKQLQYATGEIYYDSNPDF 2364  
Db 421 LYYDLOGLHFLAMESSGEYVYASDNTGTPLAVFSINGLMIKQLQYATGEIYYDSNPDF 480

Qy 2365 QMWIGFHGGLYDPLTKLVHFTQRDYDVLAGRWTSPDYTMKNVKGEPAPFNLYMFKSNPN 2424  
Db 481 QMWIGFHGGLYDPLTKLVHFTQRDYDVLAGRWTSPDYTMKNVKGEPAPFNLYMFKSNPN 540

Qy 2425 LSSELDLKNYVTDVKSMLWMFGFQLSNIIPGPFRKAKMYFVPPPYELSSQASENGOLITG 2484  
Db 541 LSSELDLKNYVTDVKSMLWMFGFQLSNIIPGPFRKAKMYFVPPPYELSSQASENGOLITG 600

Qy 2485 VQOTTERINQAFMALEGQVITKKLHASIREKAGHWFATTTPIIGKIGIMFAIKEGRVTTGV 2544  
Db 601 VQOTTERINQAFMALEGQVITKKLHASIREKAGHWFATTTPIIGKIGIMFAIKEGRVTTGV 660

Qy 2545 SSIASEDSRKVASVLLNAYYLDKMHYSIEGDKTHYFVKIGSADGLVLTGTTIGRKVL 2604  
Db 661 SSIASEDSRKVASVLLNAYYLDKMHYSIEGDKTHYFVKIGSADGLVLTGTTIGRKVL 720

Qy 2605 GNVTVSQPTLLVNGTRRRRTNIEFQYSTLLSIRYGLTPDTLDEEKARVLDQARQALG 2664  
Db 721 GNVTVSQPTLLVNGTRRRRTNIEFQYSTLLSIRYGLTPDTLDEEKARVLDQARQALG 780

Qy 2665 TAWAKEQKARDGREGSLRTEGKQQLLSTGRVQYEGYVYVPEQYPELADSSNIQF 2724  
Db 781 TAWAKEQKARDGREGSLRTEGKQQLLSTGRVQYEGYVYVPEQYPELADSSNIQF 840

Qy 2725 LRQNEGMGR 2733  
Db 841 LRQNEGMGR 849

RESULT 3  
S47008  
tenascin-like protein - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 27-Jan-1995 #sequence\_revision 27-Jan-1995 #text\_change 21-Jul-2000  
C:Accession: S47008  
R:Baumgartner, S.; Martin, D.; Hagios, C.; Chiquet-Ehrismann, R.  
EMBO J. 13, 3728-3740, 1994  
A:Title: Ten(m), a Drosophila gene related to tenascin, is a new pair-rule gene.  
A:Reference number: S47008; MUID:94349920  
A:Cross-references: EMBL:X73154; NID:g510505; PIDN:CAA51678.1; PID:g510506  
C:Genetics:  
A:Gene: FlyBase:Ten-m  
A:Cross-references: FlyBase:FBgn0004449  
C:Superfamily: unassigned EGF-related proteins; EGF homology  
F:298-324/Domain: EGF homology <EGF>  
F:394-421/Domain: EGF homology <EGF>

Query Match 26.9%; Score 3916.5; DB 2; Length 2515;  
Best Local Similarity 34.7%; Pred. No. 7.8e-200;  
Matches 836; Conservative 475; Mismatches 922; Indels 177; Gaps 52;

Qy 408 GAEVGRVTVQVPPGVFWRSOIHSIQFLEKFNISLGDALFGVYIRGLPPSHAQYDF 467  
Db 177 GOITLQGLKTRIQYISYWNMOFYOSEPAYVAFDTITPRGASIGYVGRNALPTHTQYHF 236

Qy 468 MERLDGKWKSVYSEPRRRSITQLVQNEAVEVQYLDVGLMHLAFYNDGKD-KEMVSFT 526  
Db 237 KEVLSS----FSASTRTARAHLSTIREV--TRYMEGHWFVSLYNDGDVQELTFYAA 289

Qy 527 VVLDVSQDCPRNCHNGECVSGVCHCFPGFLGADCAKACPVLCGNGOYSGTKCQCYSG 586  
Db 290 VAEDMTQNCPNGCSGNGQCLLGHCCQCNPNFGGDDGSESVCPLCSQHGHEYTNGECTCNP 349

Qy 587 WGAECDDVPMQCIDPSGGHGCSDGNVCVCSAGYKGEHCHEVDCLDPTCSHGGVGVNAGE 646  
Db 350 WKGECSLRHDEADVDCSGHGHCVSGQCQMRGKYGKFCFEEVDCPPHNCNCGHGFCA 409

QY 647 CLCSPHGGGLNC---ELARVQCPCQCGSGHGTYPDPDGLCLSCDPNWMGPDGCSVEVCSVDCG 703  
Db 410 CICKKMGKGPDCATMDQDALQCTCPDSCGSGHGTFLDPTCEAKVSGDDCKSKELCLDLCG 469  
QY 704 THGVCJGGACRCREGWTGAACDORVCHPRCIEHGTCKDGKCEGREGWNGHCHTIGROTAG 763  
Db 470 OHGRCGEGDACADPEWNGEYCNTRLCDOVRNEHQCKNGTCLCTVGWNGKHCTI----- 533  
QY 764 TETDGPDLGNGRCTL-QGNSWCVCQGTGRGPCNVMAMETSCADKNKNEGDGLVDCIL 822  
Db 524 ---EGCPNSCAGHCQCRVSGEGQWECRCYEGWDPGCGITALELNCGDSKDNKDLGVDC 580  
QY 823 DPCCLOSAQONCLLCGRSDPLDIIIOGQTDWPAV-KSFYDRIKLAGKDSHTIIPGEN 881  
Db 581 DPCCASHVCKTSQLCVSAKPTIDVLRKP--PAITASFERRMKFLIDESSLQNTAKLE 638  
QY 882 PFNSLSVLIRGOVVTDTGTPLVGVNVSVFKYPKYGTITROBGTDFDLTANGASLTLP 941  
Db 639 TPNESRAVIRGRVTVSLGMLGVVRVSTTLL-FTLTRDDGWFDMVNGGAVTLQF 697  
QY 942 ERAPFMSQERTVWLPNWSFYAMDVLNMTKE-----NSIPSCDLSGFRPDPPIIISPL 995  
Db 698 GRAPFRPQSRIVQVPNVEVVIIDLVNMSSEKGLAVTTHTTCFAHDYDLMKPVVLASWK 757  
QY 996 STEFSAAPGONPIVETQVLHBEIELPGSNVKRLYLSSRTAGYKSLKITMTOSTVPLNL 1055  
Db 758 HGQGACPOKRSALAESQVLESQIPGTGLNLVYHSSRAAGYLSIKLQLTDPDVTSL 817  
QY 1056 IRYHLMAVVEGHUFQKSFQASPLASTFIDWKTDAYQORVYGLSDAVSVGFYEYTCPSL 1115  
Db 818 HLJHLRITIGILFERFEADPGIKFTYAWNRLNIYQRYGYTTAVKVGYYTDCD- 876  
QY 1116 ILWEKRTALQGFELDPNSLGGWSLKKHILNYSKGLHKGTEGNOFLTOQPALITSIMG 1175  
Db 877 IYWDIQTTKLSGHDMSTSEVGGWNLDTIHRYNPHEGLTQKDGCSNLYLRNKPRIILLTGM 936  
QY 1176 NGRRRSISPCSCNGLAEGNKLLAPVALAVGIDGLSVGVDFNYIRRRFPNVTSLLELRN 1235  
Db 937 DGHQPLECPDQCATKORLLAPVALAAAPDGLSVGVDFNYIRRMIDGSIETVVKL-- 994  
QY 1236 KEFKHNNFAHYLLAVDPVSGSLYVSDTNSRRIRYVKLSGTLKLAGNSEVAVAGTGEOC 1295  
Db 995 ---NATRVSYRYHMAISPDLGTLVYSDPESHQIIRVROTNDYSQPELWNAEAVVSGERC 1050  
QY 1296 LPDEARCGDGGKAIDATLMSPRGIAVDKNGLMFYVDATMIRKVDQNGIISTLLGSN-DL 1354  
Db 1051 LPGDEAHCGDGALEKAKLAYPKGIAISSDNILYFADGTNIRWVDRDGIYSTLIGNHMK 1110  
QY 1355 TAVRPLSCDSSMDVAQVRLEWPTDLAVNPMDNSLYVLENNVILRITENHOVSIAGRPXH 1414  
Db 1111 SHWKIPCEBTLKLEEMHURWPELAVSPMDNTLHIIIDHMLIRMTPDGRVVRVISGRPLH 1170  
QY 1415 CQVPGIDYLSKIAIHSALESASAIASHGTGVLYIPETDEKKINRLRQVTTNGEICLLAG 1474  
Db 1171 CATASTAYD-TDLATHTLVMPQSIAGPLGELYVAESDSQRINRVITCDGRIAPFAG 1229  
QY 1475 AASDCCKNDVNCYSGDDAYATDAILNPSPLAVAPDGTIYIADLGNIRIRAVSKNP 1534  
Db 1230 AESKCNCL-ERGDCFEAEHYLATSAKFNTIAALAVTPDSHVHIADQANRYIRSVMSPIP 1288  
QY 1535 VLNAFNOQEAAAPGEOLYVFNADGTHQVTSVLTGEYLYNETY---STDNDVTELIDNN 1591  
Db 1289 EASPSREYIYAPDMQOEYIFNRFEGHVSSTRNLTGETTVFTYNTSNGKLSVTDDA 1348  
QY 1592 GNSLKIRRDSSGMPHRLLPDQNIITLVCTNGSLGWSTQNLLEGLMTYDGTGLLATK 1651  
Db 1349 GNVKVELLOYTSQVNSIENTKGOKCLRMTRMKMLHELSTPDPNYVYEHGPTGLLRK 1408  
QY 1652 SDETGWTFYDDYDHEGRLTNRTPCTGVVTSLSHREMP-KSITIDI-BNSNRD-----DDVT 1704  
Db 1409 LDSTGRSYVNVNDEFGRLTSAVTPTRGVIELSPDLISVKGAQVKSSENAQKMSLLIQGAT 1468  
QY 1705 VITNLSSVEASYTVVQDVRNSVQLNCTLRVMYANGMGSIFHSEPHVLACTIITPIGR 1764

Db 1469 VIYRNAAASRTTVDMD-----GSTTITPMGHNLQMEVAPYTTILABQSPLLGE 1517  
QY 1765 CNLSLPEW-----NGLNSTEW--LRKE-----QIKG---KVTTFGRKLVRHGR 1803  
Db 1518 ---SYPPAKQRTIEAGDLANREFWRYFYRRQOPLAQKOSKGPPTVEVGRKLVRNGD 1574  
QY 1804 NLSIDYDRNRIRTEKIYDHRKFTLRIIYDQVGRPFLMLPSSG-LAAVNVSYFNGRLAG 1862  
Db 1575 NVLTLEVDETQSVVMVMDKQELLNVYDRTSRPISFRPQSGDYAVVDLEYDRFGRLV 1634  
QY 1863 LQRCAMSERTDIDKQRIYRSMFADGKVMWSYSYLDK---SMVLLQSORQ---YTFEVDSSD 1918  
Db 1635 WKGVLOEAYSFORNGLNEIKYGDGSTMVYAFKDFMGLPLKVTTPRRSDYLLQYDDAG 1694  
QY 1919 RLLAVTMSVARHSMTHTSIGYIRMYNPPEASVIFDYSDGRILKTSFLTGROVF 1978  
Db 1695 ALOSLLTPRGHITHAFSLQSLGFFKYQYSPINRHFETILYNDEGQILAKIHPHOSKVA 1754  
QY 1979 KYGKLSKLSIEIYVDSTAVTFYDDETTGVLMKYNLQSGGFSCTIYR-KIGPLVDKQIYR 2037  
Db 1755 FVHDTAGRLLETILAGLSSTHYTYQDTSLVKSVESVEOPGPELRREFKYHAGILKDEKLR 1814  
QY 2038 FSEEGVNAFEDTYTHDNFRIASIKPVISETPLVDLRYDEISGKVEHFGFEGYIYD 2097  
Db 1815 GSKNSLASARIKAY-DGNARLSGIEAIDDKELPTTRYKYSONLGOLE-----VVQD 1866  
QY 2098 I-----NQIITAYMTLSKHF-----DTHGRKEVOYEMFRSLMYMTVOYDSMGRVIKR 2147  
Db 1867 LKTRAFNRTVQDSAKOFFAIVDYDOHGRVKSVMKNYIDVFRLELDYDLNRKISQ 1926  
QY 2148 ELKGPYANTKYTYDYDGGQLQSVAVNDRPTWRYSYDLNGN-LHLLNPGNSVRLMPLR 2206  
Db 1927 KTTFGRTAFDKKNYNADG---HVVEVLGNTNNKYLDFENGNTGVVVOGQEFN---LG 1979  
QY 2207 YLDRDRIITRLGDVQY-KIDDDGYLCORGSDIFEYNSKGLLTRYNKA--SGWSVQVRYDG 2263  
Db 1980 YDIGDRVIVKGVDFEYNNYDARGFVVRKGEQKYRNNRGLIHSFEREFQSW---YYDD 2036  
QY 2264 VGRASYKTNLGLHLYFYSDLHNPTRIITHVYHNSSETLSLYYDQGLHFLMAESSGEE 2323  
Db 2037 RSLVAVHNDKNGTQYYANPRTPHLVTHVHPFKLSRTMKLFYDDRDMLIALEHD-QR 2095  
QY 2324 YVASONTGTPLAVFINGLMIKOLOYTAYGEIYYDSNPDQFQMVIFGHGLYDPLKLVH 2383  
Db 2096 YVATDQNGSPLAFFDQNGSIVKEMKRTPGRIIKDTKPEFFVPIDFHGGLIDPHTKLVY 2155  
QY 2384 FTORDYVLAGRWTSPDYTMKNVKGK---PAPFNLYMEKSNPNLSELDLKNYVTDVKS 2440  
Db 2156 TEORQYDPHVQGMTP---LWETLATEMSHTDVFYIYHNDNPINPNKP-QNYMIDLDS 2311  
QY 2441 WLVMFGQLSNIITPGFPRAKMVEVPPPYELSESQASENGQITGVQQTTERHNAQFM--- 2497  
Db 2212 WLQFCYDLNNMOSSRYTKLAQYTPQASIKSNLAPDFG-VISGLECIKTEKSEKSDPD 2270  
QY 2498 -----ALEQVITKKLUHISIREKAGHWFATTTPIIOGKIMFAIKEGR-----VVTGVSSIA 2548  
Db 2271 FVPKPLLTKEPKMRNLLPRVSYRRG-----VFGEVLLSRIGGRALVSVVDGNSV 2322  
QY 2549 SEDSRKVASVNNAYVLDKMHYSIECKDTHYFV-----KIGSADGLVLTGTTIGRKVLE 2603  
Db 2323 QD---VSSVFNNSYFELD-LHFSIHQDQVYFVKONVKLRRDNEELRGGMFNTSTHE 2378  
QY 2604 SGVNTVVSQPTLLVNGTRRFRFTNIEFYSTLLSIRISGLTPTDLPDEKARVLDQARQAL 2663  
Db 2379 ISDHGGSAAKELRLHG-----PDADVIIKYGVDP---EOERHLKHAHKRAV 2423  
QY 2664 GTAWAKEQKARDGRGGRSLWTEGEKQQLLSTGRVOGYEGYVLPVQYPELADSSNTQ 2723  
Db 2424 ERAWELEKOLVAAGFOGRGDWTEEEKELVQHGVDGWNIGIDHSHKHYPQLADDPGNA 2483  
QY 2724 ELQRQNEWGRK 2733

Db 2484 FORDAKRRR 2493

RESULT 4

A54148

od2 protein - fruit fly (*Drosophila* sp.)

C:Species: *Drosophila* sp.

C:Date: 02-Aug-1994 #sequence\_revision 02-Aug-1994 #text\_change 11-Jan-2000

C:Accession: A54148

R:Levine, A.; Bashan-Ahrend, A.; Budai-Hadrian, O.; Gartenberg, D.; Menasheerow, S.; Wide

Cell 77, 587-598, 1994

A:Title: odd Oz: a novel *Drosophila* pair rule gene.

A:Reference number: A54148; MUID:94243925

A:Accession: A54148

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-2406 <LEV>

C:Genetics:

C:Gene: FlyBase:Ten-m

C:Cross-references: FlyBase:FBgn0004449

C:Superfamily: unassigned EGF-related proteins; EGF homology

F:514-540/Domain: EGF homology <EGF>

F:610-637/Domain: EGF homology <EGF1>

Query Match 24.9%; Score 3637.5; DB 2; Length 2406;

Best Local Similarity 32.9%; Pred. No. 5.3e-185;

Matches 817; Conservative 444; Mismatches 901; Indels 325; Gaps 59;

QY 165 DENGCRPTPTSSPLPSAQLPSHNPPVSCQPLDLSNT-SHQIMDTPDEEFSNSY 223

Db 14 DVGGPTPGHAHPH-AGRTLPMSHGRPTDGLGVHGSQTLOHQ-----NQ 60

QY 224 LLRACSGPOAAS-----SSGPP---NHHQSOT-----LRPPLPPPHN 258

Db 61 NLQAVQAASSHYDEYQHLAHPPTVNTAQRTHGROGFLLEGVTPAPDPVPRNP 120

QY 259 TLSHHSSANLNRNSTNRSSQIHAPAPA---PNDLPTSPESVQLQDSWYLSN---VPL 313

Db 121 TMSR-----MONGRLTVNPNADAFEPSCVLRTVPSGNVTPSGNLTNKGSP 168

QY 314 ETRHFLFKTSSGTPLESSSPGYPLTSGVYTP--PPRL-----LPNTRSRKAFKKLP 367

Db 169 D-----FKSGSACTPTKDTLKGYERTQCGMGVLPQRSVMNGLPAHHYSAPMFRKDL 223

QY 368 SKYCS---WKCAALSAIAAALLAILLAYE-IVPWS----- 399

Db 224 VARCSPWFSGISVLFVAFVWMLILTTTGVTKWNSPPCSVLVNGEASEVTAAKSTNT 283

QY 400 ----LNKSSI-----DSGEAE----- 411

Db 284 DLSKLNSSVRKNGOGIGLAQSGGLGAAGVGGGGSSAATVTTATNSGTAGQLQSTS 343

QY 412 -----VGRVTO 418

Db 344 ASAEATSAATSSQSLSLTSLSSLANANGARTPAKSPFDPDGTFTGQILGQKLTK 403

QY 419 EVPPGVFWRSQIHISQPFKFNISLGLKDALFGVYIRRGLPSPSHAQYDFMERLDGKEKWS 478

Db 404 EIOPYSYNNQFYQSEPAYVKFDYTPRGASIGYGRNALPTHQYHFEVLG----- 458

QY 479 VVESPERRSTIQLVQNEAVFVQYLDVGLWHLAFYNDGKD-KEMVSFNTVVVLSVQDCPR 537

Db 459 FSASTRTARAHLSTIREV--TRYMEPGHFWLSLYNDGDVQELTFYAAVAEDMTQNCN 516

QY 538 NCHNGECVSGVCHCFPGFLGACAKAACPVLCSGNGQYKSGTCQYSGWKGAECVPMN 597

Db 517 GCSGQGLLGHQCNQPFQGGDDCSQVPLCSQHGHEYNGECINPQGWKKECSLRHD 576

QY 598 QCIDPSCGGHSGCIDGNCVCSAGYKGEHCHEVDCLDPTCSHGVYCVNGECLCSPGWGLN 657

Db 577 ECEVADCSGHGCHVSGKQCQMRGKYGKFCBEVDCPHPCNSGHGFCADGTCICKKGWGP 636

QY 658 C---ELARVQPCQCSGHGTYLPDPTGLCSDPNWMPGDCSVEVCSVDCTGHGVCIGACR 714

Db 637 CATMDQDALQCLPDCSGHGTFLDTQTCTCEAKWSGDCSKELCDLDCGQHGRCGDACA 696

QY 715 CEEGWTGAACDQQRVCHPRCIEHGTCTDGKCECEGNGEHCITIGRTAGTETDCCPLCN 774

Db 697 CDEWGEYCNTRLCDVRCNEHGCKNGTCLCVTGNWGHCTI-----EGCPNSCA 747

QY 775 GNGRCTL-GONSQCVCOTGRPGCNVAMETSCADNKNKGDLVCLDPLDCCLOSACO 833

Db 748 GHGQCRVSGGWECEYEGWDPDGCIALLELNCGDKNKDKGLVDCDEPECCASHVCK 807

QY 834 NSLLCRGRDPLDIIQOQTDWPAV-KSFYDRIKLLAGKDSHTIIPGENPNSLSVLIR 892

Db 808 TSOLCVSAPKPIDVLLRKQP--PAITASFFERMKFLIDESSLQNYAKLETFNESRAVIR 865

QY 893 GOVVTDDCTPLGVNVSFVKYKYTYITRODGTFLDIANGASLTLLHFERAPMSERT 952

Db 866 GRVYTSLGMGLVGRVSTTTTLE-GFTLTTRDDGWFDMVNGGAVTLQFGRAPFQPSRI 924

QY 953 VLPWNSFYAMDITLVKTEB-----NSIPSCDLSGVRPDPPIISSPLSTFFESAAPQN 1006

Db 925 VQVPWNEVVIIDLVVMSMSEKGLAVTTHTCFADYDLMKPVVLAWSKKHGFQACFDRS 984

QY 1007 PIVPETQVLHEIEELPGSNVLYLRSRTAGYKSLKLTMTQSTVPLNLRVHLMVAEG 1066

Db 985 AILAESQVIOESLQIPGTGLNVVHSSRAAGYLSIKLQLTDPDVPITSLHLIRITIEG 1044

QY 1067 HLQKFSQASPNLASTPDKDAYGQRYVGLSDAVVSGVEYETCPSLILWEKRTALLO 1126

Db 1045 ILFERIFEADPGKFTYAWNRNIYRQRYGVTTAVVKGVQYVTDCTD-IVMDIQTTLKS 1103

QY 1127 GFELDPNGLGWSLKHILNVKSGILHKGTFGENQFTQOPAIITISMGRRRSISCPSS 1186

Db 1104 GHDMSEVGGWNLDIHRYNFHEGILQKGDGNSYILRNKPRILLTMTGDSGHQRPCEPD 1163

QY 1187 CNGLAENKLLAPVALADIGSLYGVDFNYIRRFPSNRVNTSLELRNKEFKHNNPAH 1246

Db 1164 CDGQATKORLLAPVALAAPDGLFVDFNYIRRMIDGSIPTVVKL-----NATRVSY 1217

QY 1247 KYVLAVDPVSGSLVSDTNSRRYRVKLSGTDNACNSEVAVAGTGEQCLPFDGARGDG 1306

Db 1218 RYHMAUSPLDGLTVSDPESHQIIRVTDNDYSQPELWNAVVGSGERCLPDDGARGCGD 1277

QY 1307 GKATDAPLMSPRGAVDKNGLMFVDATMIRKVDQNGIITLLGSN-DLTAVRPLSCDSS 1365

Db 1278 ALAKADAKLAPKGAISSDNLYFADCTNIRWVDRDGIIVTLGNHMHKSHWKPIPEGT 1337

QY 1366 MDVAQVBLEWPTDLAVNPMONSXVLENNVILRTITENHQVSIAGRPMHCOVPOIDYSL 1425

Db 1338 LKLEEMHLRWPTELAVSPMDNTLHIIIDHMLRMTDGRVIRVISRPLHCAATASTAYD-T 1396

QY 1426 KLATHSALESASATAISHTGVLYITETDEKKINRLROVTTNGEICLLAGASDCCKNDV 1485

Db 1397 DLATHATLVWPOSATFGPLGELYVAESDSQRIKRVIRVITGDRIAPPAGAEKNCNCL-ER 1455

QY 1486 NCNYSGDDAYATDAILNPSLSLAVAPDGTIYIADLNIRIRAYSKNPKVNLNAPNOYEA 1545

Db 1456 GDCDFEAHVLATSAKENTIAALAVTPDSHVHIAADQANYRIRSVMSIPEASPSREVEIY 1515

QY 1546 SPGEQELVFNADGIIHOYTSLVTEGYLYNFTY---STDNDVTELIDNNGSLKIRDS 1602

Db 1516 APDMQEIYIFNRFQHVSTRNLTGETTYVFTYVNTSNGKLSVTDAAGNKKVLLRDY 1575

QY 1603 GMPHLLMPDQIITLVGTNGGLKVVSTONLEGLMTYDGTGLTATKSDGTGWTTFYD 1662

Db 1576 SQVNSIENTGKQCRRLRTRMKMLHELSTPDNYNVTYEGHTGLLRTKLDSTGRSVYN 1635

QY 1663 YDEGRITNVTPTGVTSLHREME-KSITIDI-ENSNRD-----DDVTITNLSSEAS 1715

Db 1636 YDEGRITSAVTPGRIEVSFDSLVSVKGAQVKVSENQAQKMSLLIOGATVIVRNGAESR 1695

QY 1716 YTVQDQVRNSYQLCNNGTLRVMYANGMGTISFHSHPHVLAGTITPTIGRCNISLPM- 1772

Db 1696 TTVDMD-----GTTTTPWGHNLQMEVAPYTTILAEQSPLGE---SYVPAPAK 1741  
Qy 1773 -----NGLNSIEWR--LRKE-----QIKG---KVTIFGRKLRVHGRNLLSIDYDRNI 1814  
Db 1742 RTEIAGDLANREWFYFVRQOPLQAGSKGPPRPVTEVGRKLRVNGDNVLTLEYDRET 1801  
Qy 1815 RTEKIYDHRKFTLRIYDQGRPLWLPSSG-LAAVNVSYFFNGRLAGLQRGAMSERTD 1873  
Db 1802 QSVWVMDKQELLNVYDRTSRPISFRPQSGDYADVLDLEYDRFGLVSKWGLQEAYS 1861  
Qy 1874 IDKGRIVSRMADGKVNYSYLDK--SNVLLLSQORQ--YIFEYDSSDRLLAVTMPVA 1929  
Db 1862 FDRNRLNIEIKYDGGSTWYAFKMGFSLPLKVTYPPRRSDYLLQYDDAGALQSLTPRGH 1921  
Qy 1930 RHSMSTHSIGYIRNTYNPPESNASVIFDYSDGRLTKTSFLTGROVPYKYGKLSKSE 1989  
Db 1922 IHAFLSLOTSLGFFKYQYVSPINRHPFEILYNDGQILAKIHPHQSKVAFVHDTAGRLET 1981  
Qy 1990 IYDSTAVTFGYDETGVLLKMYNLQSGFSCITRYR-KIGPLVDKQIYRFSEGMVNARF 2048  
Db 1982 ILAGLSSTHYTQDTSLSKVSVEQEPGELRREFKYHAGILKDEKLRFGSKNSLASARY 2041  
Qy 2049 DTYHDNSFRIASIKRPVISETLPVDLYRYDEISGKVEHFGFEGYIYDI----NQIITT 2104  
Db 2042 KYAY-DGNARLSGIEAIDDKELPTTRYKYSONLGOE-----VVQDLKTIIRNAFNRT 2093  
Qy 2105 AVMTLSKHP-----DTHGRKEVOYEMFRSLMYNWTVOYDSMGRVIRKELKLPYANTT 2158  
Db 2094 VTQDSCKQFFAIVDYDQGRVKSVMKNYKIDVFRLELDYDLNRKLSKQTTFFGRSTAFD 2153  
Qy 2159 KYTYDYGQGLQSVAVNDRTWRYSYDLNGN-LHLLNPGNSVRLMPLRYDLRDRITRLG 2217  
Db 2154 KINYNADG---HVEVLGNTNKKYLFDENGNTGVVDOQGEKFN---LGYDIDGRVIVG 2206  
Qy 2218 DVQY-KIDDGVLQCGSGDIFEYNSKGLLTRYANKA--SGWSVOYRYDGVGRASYKTNL 2274  
Db 2207 DVEFNNDYDARGFVVRKGEQRYNNRGLIHSFERFOSW---YYDDRSRLVAMHDNK 2263  
Qy 2275 GHHLQYFSDLNPRTHIVYNNHNSSETLSLYDQLQGLHFAMESSGEEYVASDNTGTP 2334  
Db 2264 GNTQYVYANPRTPHLVTHVHPFKISRTMKLFYDDRDMLIALEHED-QRYVYVTTDQNGSP 2322  
Qy 2335 LAVFSINGLMIKOLQYATGEIYYSNPDQFQWVGFHGLYDPLKLVHFTORDYDLVLAG 2394  
Db 2323 LAFFDQNGSIVKEMKRTPFGRITKOTKPEFFPIDFHGGLIDPHTKLVYTEQORYDPHVG 2382  
Qy 2395 RWTSPDYTMKNVKGKEPA-PENLYMFK 2420  
Db 2383 QWMTTP---LWETLATEMSHTPDVFYIR 2406

RESULT 5  
Ti6743

hypothetical protein R13F6.4 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Mar-2000

C:Accession: Ti6743

R:Miller, N.

submitted to the EMBL Data Library, April 1994

A:Description: The sequence of *C. elegans* cosmid R13F6.

A:Reference number: 218570

A:Accession: Ti6743

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2531 <MTL>

A:Cross-references: ENBL:U00046; NID:q470358; PID:q470361; PIDN:AAC47044.1; GSPDB:GN0002

A:Experimental source: strain Bristol N2; clone R13F6

C:Genetics:

A:Gene: CESP:R13F6.4

A:Map position: 3

A:Introns: 52/2; 113/1; 170/2; 594/1; 1009/1; 1181/1; 1547/3; 1672/3; 1890/1; 2278/1; 24

C:Superfamily: unassigned EGF-related proteins; EGF homology

Query Match 15.1%; Score 2208.5; DB 2; Length 2531;  
Best Local Similarity 25.5%; Pred. No. 8.3e-109;  
Matches 694; Conservative 441; Mismatches 1008; Indels 581; Gaps 99;

Qy 322 TSSGS-TPLFSSSP---GYPLTSGT-----VYPPRLPLRNTFSRK 360  
Db 22 TSGGAPNPTYSDASTLLKYPLAAGTNQNRQRQVGTMMNGDPVAGGPMAL-----SKK 75  
Qy 361 AKFLKKPKYCS-WKCAALSAATAAALLA---ILLAYFIVPWSLKNSSIDSGEA-----410  
Db 76 KKKFDDSDTCSRWPWSKWNILLAAALLVALFICILLFRAPNYVYTPAPASSDSSAAA 135  
Qy 411 -----EVRRVTQEVPPGVFWRSGIHSIQPQFLKFNLSLCKDALFGVY 453  
Db 136 AASRYQDLGLRALPAISLGERVDVEFFPKSMATTELTVTKPSRIFRNATVSGAQLVLL 195  
Qy 454 IRRGLPPSHAOYDFM-----ERL-DGKEKWSVVE-----SPRRRSIOTLVON 495  
Db 196 MSAGVHPSLSLHDALFPADRIRDSKSPTHIVEEFGSRRSRSLGASSSRHRNIEILSPR 255  
Qy 496 EAVFVQYLDVGLHLAFYNDGDKEMVSENTVVD-----SVQ 533  
Db 256 SATFEQFVLEGRHYLTFINERSRVEPISFVAEELQRPTTPTSSSGTGAKEHPLASVL 315  
Qy 534 DCPNRCHNGECVSGVCHCFPGFLGADCAKAAACPVLCGSGQYSGKTQCYSGWKGAEC 593  
Db 316 VCESNCNQRGECVHGKCHCAPGFTGTCDEAVCPVCSGNGVFGSGICVCKSGFKGKE 375  
Qy 594 VPMNCIDPSCGGHSGC-IDGNCVCSAGYKGHBCEVDCDLPDTCSSHGVGVNCECLSPG 652  
Db 376 MHRNCEVADCNRGRCDDTGRRCNPGWTGEACELRPHASCHDRGVGVNGTCYCMDG 435  
Qy 653 WGLANCE-----L 660  
Db 436 WRGNDCSVPADAIHVHVPQNSPPRRGQEPTESSKTRKAQVKPTPTSEKKESRELOKPII 495  
Qy 661 ARVQCPDQ----CSHGTYLPDTGLCSDPNMGPDCSVCEVSGVDCGTHGVCI-GGACRC 715  
Db 496 ATVQPTTESSHPCSAHGQLIDD--ICQESGWDVDCSQAC--QC-VNGDCLDDGSCQC 550  
Qy 716 EBGWTAACDQVCHPRCIEHGTG-KDGCECRBEGWHECHTIGRTAGTETDGCPLCN 774  
Db 551 WKGWRGSCNCTDKKAIGCEDRGKASDGSCSKCSGNGENCAI-----DGCNQC 601  
Qy 775 GNGRCTLGONS--WQCVQGTWRGPGCNVAMETSCADNKNDEGDLVDCDPCCLQASAC 832  
Db 602 KGCEGMDRRSSEWSCRCQAGSTGYDCSVSVEMHCDGDLNDSGLIDCDDECCSSSC 661  
Qy 833 QNSLLCRSRDPLDIIQOQGTDMPAV--KSFYDRIKLAGKDSHTIIPGENPNSLSVL 890  
Db 662 SSESVCSTAASPIELMR----MPPIFANFAQRVGLFIMEKSVQSYTDSQFENENLISV 717  
Qy 891 IRGVVV-----TTDGTPLGVNVSVFKYKYTYTITRODGTDLANGSA 935  
Db 718 INGRVWGGSGPTGSDDLSTYSNKSVPVGVVSDAAHPLYGTLTREDGYFDLVNGAR 777  
Qy 936 SLTLHERAPFMSQERTVWLPWNSFYAMDTLVMTKEENSIP-----SCDLSGFVR 985  
Db 778 SVTLQFLRTQSVKSKSVFSPROIHDDIVLYQSGSPPAISWAPARAKCSPLRRI 837  
Qy 986 PDPIIISPLSTF-----FSAAPGQNPVETQVLHEELPGSNVKLYLSRRTAGYKSL 1041  
Db 838 PDVVLISNMQVTSDGIETDETSDSRIVVDSRISFESLPQGTDLVRLVYDSARSAPST 897  
Qy 1042 LKITWTQSTVPLNLRVHLMAVEGHLFKQSFQASPNLASTFIWDKTDAYGORVYGLSDA 1101  
Db 898 MLIGLLYDRVDKELRKVHINIRIAGRFRDVLAPRTNITYVFAWDMKNAYROSESGLPV 957  
Qy 1102 VVSVGEFEYETC--PSLLILWEKRTALLQGFELDPNSLNGWSLDKHHILNVKSGILHKGTGE 1159  
Db 958 TVRVGYEQCDRTSERVWQTRRSOMMGATARKMIGTMTWTLDIHHLLDIVNVNVEGNGG 1017



Qy	335	PGYPL-TSGVTYTPPPELLPRNTFTSRKAPLKKPSKY-CSWKAALSAIAAALLAILLA	392
Dd	2	PGPELRNNSAHAP-----HYSPISPRSHIDRKCHRCWRCLSTIALIFVSVLTAMLA	56
Qy	393	YFTVPWSLK-----	401
Dd	57	YFAVSMSKNMDSNCILLVODVKSQPHDLHGGLAKSNEKGVAATPTTEESTQTSTSDHG	116
Qy	402	-----NSSIDSEAEVGRR---VTQEVF-	421
Dd	117	QNGHGLMNPAGSGGNSGTQQOOLLVQOOOPHSINQPLTPLDATNTHLDHHOLTYGGAL	176
Qy	422	-----PG-----	423
Dd	177	PGGVGGTGMGGMMNGGLNGLGGLMQOPGGGLNGHHQAQLPOLGCVGVVELKEFNAY	236
Qy	424	-----VFWRSHIISQOFKLKFENLSLGDALFGVYIRRGLPSPSHAQYDFME-----RL	471
Dd	237	HATIPAVQFWTLFRNKHPAFIREFTLPGAHAFAVSRNPVAPTQHDFVEFIKGRRL	296
Qy	472	DGREK-----WSVESPRER-----RSQTLLVQNEAVF-----	499
Dd	297	DSLHRHRRSSANPPQLGKIWSCSIARTTRVRNRLRIETRNTRFVASIPANTMMWRHPSI	356
Qy	500	-----VOYLDVGLWHLIAFYNDGDKD	519
Dd	357	RPNINCINGLASPKVHVHNKRASAGDGLPALVDANTVNVSLLOYLDTGLWFVISVYND----	412
Qy	520	EMVSFNTVWLDSVD-----CPNRCHNGEGCSYGVCHEFPGLGADCAKACPVLCSGNG	574
Dd	413	ELVAHSVLLAAEAEGYSTTCPNDCSGRGSCYLGHKDCIDGYQGVDCSKSCPVLCSAHG	472
Qy	575	OYSGKTGCOCYSGHKGAECDPVMNQCIDPCGGHSGCIGDNCCVCSAGYGEHCREEVDCILD	634
Dd	473	HYGGGVCHCEBEGWKGAECDDIPVECEVPCNSSSHGRICEGECBERGWGPYCQDHQCILD	532
Qy	635	TCSHHGVYNGECLCSPGWGGLNC----ELARYQCPCQCSGHGTYLPDTGLACSDPDNMNGP	691
Dd	533	LCSGHGTCAVAGOCYCRAGWGEDCGTIIDQVYQCLPGCSEHGYTDLETGCQCVCEHHWTGP	592
Qy	692	DCSVEVCSVDCGTHGVCIGGACRCEBGWTGAACDQRVCHPRIEHGTCCKDGKCEBREGWN	751
Dd	593	DCSOAVCSLDCGRNGVYESGKCRCNSGWTGNLCDQLPCDSCRSEHQCKNCTCVCSGOWN	652
Qy	752	GEHTCTGRQTAGTETDCCPLDCNNGRCRTLQONSWCVCOTCWGRPCCGNVAMEYSCADNK	811
Dd	653	GRCTL-----PGCENGCSRHOQCTLENGEYFCDCIEGWAAGSDCSIALELNCKDNI	703
Qy	812	DNEGDELVDCLDPDCCLOQSACONSLLCRGRPLDIIOOGQTDMPAV-KSFYDRIKLL	868
Dd	704	DNDGDGTTCDSDCECSHPACSEHIINCLSSNDPVEYLLRKOB--PSVTAASFVORVKFL	759

	Query Match	5.08;	Score 726;	DB 1;	Length 1810;	
	Best Local Similarity	21.4%;	Pred. No. 4.8e-30;			
	Matches 422;	Conservative 241;	Mismatches 715;	Indels 594;	Gaps 98;	
Qy	535	CPRNCHNGNECVSGVCHCFPGFLGADCAKACAPVLCSNGNGYSGKTQCYSGWKGABCDV	594			
		:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :				
Dd	223	CPSDNDGGKVDGVGVCFCFEYTGPDGEEELCPHGCGIHRGCVGRGCVCHEGFGTGDENE	282			
		:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :				
Qy	595	PMNCIDPSCGGHSCIDGNVCVSAGYKGRHCEEVDCLDPTCSSHGVCNVNGECLSCPWG	654			
		:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :				
Dd	283	PL--CPN-NCHNRGRVDNECVDEGYTGEDCGLICPN-DCFPDRGCIINGTCFCEGYT	338			
		:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :				
Qy	655	GLNCELARVCPODGSHTGLYPDTGLCSDPNNMGDPDCSVEGCVSDGTHGVCIIGSAGR	714			
		:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :				
Dd	339	GEDC--GELTCPNMCNGRC--ENGLCVCHEGVFDGDSQKRCKPKTCNNRGRVCDGRCV	394			
		:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :				
Qy	715	CEEGWTGAACDQ----RVCHPR-----CTEHCTCKDGK	743			
		:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :				
Dd	395	CHEGYLGEDCGELRCPNDCNHRGRCIQCVCDEGFIGDCGELRCPNDCOORGRCIHQ	454			

A:Molecule type: mRNA  
A:Residues: 1-1044,1318-1810 <J02>  
A:Cross-references: GB:J04519  
R:Spring, J.; Beck, K.; Chiquet-Ehrismann, R.  
Cell 59, 325-334, 1989  
A:Title: Two contrary functions of tenascin: dissection of the active sites by recombinant proteins  
A:Reference number: A33379; MUID:90030407  
A:Accession: A33379  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-204, 'G', 206-221, 'A', 223-380, 'D', 382-386, 'H', 388-444, 'HN', 447-450, 'V', 45  
A:Cross-references: GB:W23121; NID:9212746; PIDN:AAA49086.1; PID:9212749  
A:Accession: B33379  
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-204, 'G', 206-221, 'A', 223-380, 'D', 382-386, 'H', 388-444, 'HN', 447-450, 'V', 45  
A:Accession: C33379  
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-204, 'G', 206-221, 'A', 223-380, 'D', 382-386, 'H', 388-444, 'HN', 447-450, 'V', 45  
A:Cross-references: GB:W23121  
R:Pearson, C.A.; Pearson, D.; Shlbahara, S.; Hofsteenge, J.; Chlquet-Ehrismann, R.  
EMBO J. 7, 2977-2982, 1988  
A:Title: Tenascin: cDNA cloning and induction by TGF-beta.  
A:Reference number: S01292; MUID:89030589  
A:Accession: S01292  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 27-181, 'R', 183-204, 'G', 206-221, 'A', 223-380, 'D', 382-386, 'H', 388-444, 'HN', 447-450, 'V', 45  
A:Cross-references: EMBL:X08030

QY 744 CECREGWNGEHCITIGRGTACTETDCCPDLCNNGRCRTLGQNSWCVCQTGHRGPGCNVAM 803  
Db 455 CECEHGFIGEDGELR-----CPDNCNHCRCVNG-----QCVCDEGYTGDCG--- 498  
QY 804 ETSACADKNKNEGDLVCLDPCDCLQACQNSLL---CRSRRDPLDILIOQOQTDWPAVKS 860  
Db 499 ELRCPNDCHNRG---RCVBGR---VCONGFMGDCGELSCPCNDCHQH--- 541  
QY 861 FYDIRIKLLAGKDSPIH-IPGEN-----PENSLSVS-LIRGOVYTTDGTPLGVNVSVFK 912  
Db 542 ----RCVDGRVCYCHGFTGEDCRERSCPNDCNNVGRVCEGRVCBEG--YMGIDCSDVS 594  
QY 913 YPKYGYTIR-ODQTFDL-----IANG-----ASLTLHER 943  
Db 595 -PPTGLVTNTVTKTVNLKHEHNLVNEYLVYVPTSSGGDLDLQTFVPGNQTSATIH-EL 652  
QY 944 APFNSQERTVWLPNWSFYANDTLVMKTEENSIPSCDLSGFRVDPDPIIISPLSTFFSNAAP 1003  
653 EPGVE-----YFRVFAILNKKKSIP-----VSARVATILPAPE 686  
QY 1004 G-QNPVIVPETQVLHE----EIELPGSNVKRLYLSSR-----TAGYKSLKLTITQSTVP 1052  
Db 687 GLKFKSVRETSVQVEMDPLSIFSDGWELVFRNMOKDONGDITSSLRKRPETSYMQPGAP 746  
QY 1053 LNLIRVHLMVAVEGHLFQKSFQASPNLASTFIWDKTDAYGO-RVYGLSDAVSVGFYET 1111  
Db 747 GOQYNVSLHIV-----KNTRGPGL-SRVITTKLDAPSQIEAKDVTDTT----- 789  
QY 1112 CPSLILMEKRTALLOQFEL--DPSNLGGWSLDKHHILNVKSGILHKGTEGNOFLTQOPAI 1169  
Db 790 --ALITWSKPLAETIEGIELTYGPKDVPG---DRTTI-----DLSEDENOY----- 829  
QY 1170 ITSIMGNRRRS-----ISCPSCNGLAEGNKLLAPVALAVAGIDGSLVYGFNFYIRRPFS 1224  
Db 830 ---SIGNLRPHTEYEVETILSRGDMESDPA-----KEVFVTDLD-----AP 869  
QY 1225 RNVTSLLELRNK----EPKHSNNPAHKYLLAVDPVSGSLVSDTNRRIYRVKLSGTKDL 1281  
Db 870 RNLKRVSDTNSITLENKFSHANIDNRIKFPATSG-----PTEYIGIGVTAVRQ-DR 945  
QY 1282 AGNSEVAVAGTEOCLPDEARCQGGGKAIDATLMSPRGIAVDKNGLMYFVDATMIRKVDQ 1341  
Db 906 GDHTELVTPKGNQA-----TTRATLTGLR-----PGTEYIGIGVTAVRQ-DR 945  
QY 1342 NGIISTILGSDNLTAVRPLSCDSDMDVAQVRLMPTOLA-----VN 1382  
946 ESAPATINAGTDLNPRDLEVSDPTEIT-LSLRMRPVAKFDRLTYVSPSGKKNMEI 1004  
QY 1383 PMDNSLVLENNVILRITENHOVSIAGRPCHQVPGIDYSLSKLATHSALESATAIS 1442  
Db 1005 PVDSTSFILRG---LDAGTEYITSLVAEKGRHKSAP----- 1037  
QY 1443 HTGVLYITEDEKKINRLRQVTTNGEICLLAGASDCCKNDVNCVSGDDAVATDAIL 1502  
Db 1038 -TTIKGSTE-EEPELGNLSVSETGWDFGLWTA--DCAVENFVIQ 1080  
QY 1503 NSPSSLAVAPDGTIYIADLG---NIRIRAVSKNRPVLNAFNQYEAASPG-----EQELY 1553  
Db 1081 VQOSD---NPETWNITVPGQHSVNVYGLKANTP-----YVNTLYGVIRGYRTPLY 1130  
QY 1554 VFNADGTH-----QYTVSLVTGEYLYNFTYSTDND-----VTELDNNGNSLKIIRDSGSM 1604  
Db 1131 VETTTGAHPEVGETVSDITPE-SFNLSWTTTNGDFOAFTTEIIDSNRLLPEMFENISGN 1189  
QY 1605 PR--HL--LMPDNOIILTVCTNGLKLVVSTQNLELGLMT-----YDGNLTGLLAKSDET 1655  
Db 1190 SRTAHISGLSPSTDFIVLYIGISHGFR---TQAISAATTEAEPEVD---NLLVSDATPD 1243  
QY 1656 GWTTFDYDHEGRLTNTVTRTPGV---VTSLHREWEKS--TTIDENSNRDDDDVTIINL 1709  
Db 1244 GF-----RLT-WTADGCVDFDSFVLKIRDTKRKSDPLELIVPGHERHTDITGKEG 1292  
QY 1710 SSVSEASYTVVQDOVRNSYQLCNNGTLRVMYANGMGISFHSHPHLACTITPTIGRCNISL 1769

Db 1293 TEYEILYGVSSG-RRSQPI--NSVATTVVGSPKGISF-----SDITENSARVSWT- 1340  
QY 1770 PMENGLANSIEWRLKEOIKG---KVTIFGRKLRVHGRNLL-STDYORNIETKIIDHR 1824  
Db 1341 PPRSVD--YRVSVYVITGGTPNVVDSKTRTKLVKLVGVDYNNVILSVKGFESE 1398  
QY 1825 KFTLRIIYDOVGRPFLWLPSSGLAAVNVSYFPNGLAGLQARGASERTDIDQGRIVSRM 1884  
Db 1399 -----PISGIL-----KTALDSPSGLVVM 1418  
QY 1885 FADCK---VMSYSYLDKSMWLLLSQSQRYFEYDSSDRLLAVTMPSVARHSMSTHTSGY 1941  
Db 1419 ITDSEALATWQPA-----IAADVNIYVSYSEDE-----PEVTQ--MVGNTVEY 1461  
QY 1942 IRNIYNPESNASVIEDYSD--DGRILKTSFLGTGRQVYKYKGLSKLSELYVSDATVF 1999  
Db 1462 DLNGLRPATEYTLRVHAKVDAQKSETLSTQFT-TGLDA----PKDLSATEVQSEATVITW 1516  
QY 2000 -----GYDETTGVLLKMYNLQSGGFSCTIRYRKIGPLVDKQIYRFSEEGVWNA 2047  
Db 1517 RPPRAPVTDLTYESIDGRVKVELDPETTSYTL--TELSPST----- 1558  
QY 2048 FDIYTH----DNSPRIASIKPVISETPLVDLYRYDEISGKVEHFGKFCGVIIYDINOIIT 2103  
Db 1559 -QYTVKQLALSRSRSMKIQTFTTGL---LYPYPKDCSOAL-----LNGEVT 1603  
QY 2104 TAVMTLSKHFDTHGRKEQVEMYFERSLMYMTVOYDSGRVI---KRELKLGPAVANTTKY 2160  
Db 1604 SGLTYIYLGND-----RTQLOVCDMAEDGGWIVFLRRQNGKEDFRNWKNY 1652  
QY 2161 TYDVGDGQLOSVAVNDRPTWRYSYDLNGLNHLNPGNSVRLMLPLRYDLDRITRLGSDVQ 2220  
Db 1653 VAGF-GD-----PKDEFWIGLE-NLHKI---SSOGQYELRVLDRD---GETA 1692  
QY 2221 YKIDD-----DGYLCQRGSDIFEYNSKGLLPRAYNKASGWSVQVRYDVG 2265  
Db 1693 YAVYDKFSVGDAKTRYRLRVGDYSGTAGDSMTYHNGRSFST-----FDKDN 1738  
QY 2266 RRASYKNLGHHLHLYFYSDLHNPTRITHVYHNSSEITSLYDYLQGHLFAME 2317  
Db 1739 DSAITNCALSYKGAFWYKNCNR-VNLMGRYGDNNHNSQGVNFWHKGHEYSIQ 1789  
  
RESULT 8  
A32160  
tenascin-C - human  
N:Alternate names: hexabrachion  
C:Species: Homo sapiens (man)  
C:Date: 31-Jul-1989 #sequence\_revision 12-Apr-1996 #text\_change 20-Aug-1999  
C:Accession: I38337; A32160; S14015; S16166; S50208; S49354  
R:Gherl, R.; Carnemolla, B.; Siri, A.; Ponassi, M.; Balza, E.; Zardi, L.  
J. Biol. Chem. 270, 3429-3434, 1995  
A:Title: Human tenascin gene. Structure of the 5'-region, identification, and character  
A:Reference number: A55974; MUID:95155442  
A:Accession: I38337  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-2201 <RES>  
A:Cross-references: EMBL:X78565; NID:g556844; PIDN:CAA55309.1; PID:g556845  
R:Gulcher, J.R.; Nies, D.E.; Marton, L.S.; Stefansson, K.  
Proc. Natl. Acad. Sci. U.S.A. 86, 1588-1592, 1989  
A:Title: An alternatively spliced region of the human hexabrachion contains a repeat  
A:Reference number: A32160; MUID:89160821  
A:Accession: A32160  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 431-538, 'R', 540-1065, 'H', 1067-1599, 'LWLHPRASN', 1609-2054, 'L' <GUL>  
A:Cross-references: GB:M24630; NID:g514363; PIDN:AAA52703.1; PID:g553348  
R:Siri, A.; Carnemolla, B.; Saginati, M.; Leprini, A.; Casari, G.; Baralle, F.; Zardi  
Nucleic Acids Res. 19, 525-531, 1991  
A:Title: Human tenascin: primary structure, pre-mRNA splicing patterns and localization  
A:Reference number: S14015; MUID:91187670



A:Accession: S14015  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-243,245-369,'V',371-679,'R',681-1676,'I',1678-2139,'TR',2142,2144-2201 <SF>  
A:Cross-references: EMBL:X56160; NID:g37226; PIDN:CAA39628.1; PID:g37227  
R:Nies, D.E.; Hemesath, T.J.; Kim, J.H.; Gulcher, J.R.; Stefansson, K.  
J. Biol. Chem. 266, 2818-2823, 1991  
A:Title: The complete cDNA sequence of human hexabrachion (tenascin). A multidomain protein  
A:Reference number: S16166; MUID:91131572  
A:Accession: S16166  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-538,'R',540-1065,'H',1067-1599,'LWLHPRASN',1609-2054,'LH',2055-2201 <NIE>  
A:Cross-references: EMBL:M55618; NID:g184483; PIDN:AAA88083.1; PID:g184484  
R:Glumoff, V.; Savontaus, M.; Vehanen, J.; Vuorio, E.  
Biochim. Biophys. Acta 1219, 613-622, 1994  
A:Title: Analysis of aggrecan and tenascin gene expression in mouse skeletal tissues by  
A:Reference number: S50206; MUID:95035091  
A:Accession: S50208  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
Residues: 46-125 <GLU>  
A:Cross-references: EMBL:X80280  
Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994  
C:Genetics:  
A:Gene: GDB:HXB  
A:Cross-references: GDB:l20073; OMIM:l87380  
A:Map position: 9q33-qq33  
C:Superfamily: tenascin; EGF homology; fibrinogen beta/gamma homology; fibronectin type  
C:Keywords: alternative splicing; extracellular matrix  
F:408-434/Domain: EGF homology <EGF>  
F:622-703/Domain: fibronectin type III repeat homology <FN3A>  
F:711-794/Domain: fibronectin type III repeat homology <FN3B>  
F:892-976/Domain: fibronectin type III repeat homology <FN3C>  
F:1437-1519/Domain: fibronectin type III repeat homology <FN3D>  
F:1619-1701/Domain: fibronectin type III repeat homology <FN3E>  
F:1709-1790/Domain: fibronectin type III repeat homology <FN3F>  
F:1798-1878/Domain: fibronectin type III repeat homology <FN3G>  
F:1886-1966/Domain: fibronectin type III repeat homology <FN3H>  
F:1981-2189/Domain: fibrinogen beta/gamma homology <FEG>

[illegible]



Qy 1783 -RKQIKGK-----VTIFGRKLRVHGRNL-----LSIDYDRN----- 1813  
Db 1490 ERTAHISGLPPSTDFIVYLSGLAPSIRKTKTSATATTEALPLENLNIS-DINPYGFTVS 1548  
Qy 1814 -IRTEKIYDD-----HKKFTLRIIYDOVGRPFLWPSGLAAV 1850  
Db 1549 WMASENAFDSFLVTVVDSGKLLDPQETLSGTORKLELRGLITGIG-----YEV 1597  
Qy 1851 NVSYFFNGRLAGLQGMASERTDIDKQGRVSRMFADGKVMSYS-----YLDKSMVLLQOS 1906  
Db 1598 MVSFGTGOHQTKPLRAEIVTAEPEVDNLLVSDATPDGFRLSWTADGCVDFNFKLRTDT 1657  
Qy 1907 QRQYIFEYDSDRLAVTMPVSARHSMSTHTSIGYIRNIYNPPESNASVIFDYSDDGRIL 1966  
Db 1658 KKQ-----SEPLEITLLAPERTDLGLREATETEYELYGISKGRRSQTVS-----AI 1705  
Qy 1967 KTSPLGTGRQVFFYKGLSKLSEIYDSTAVTFGYDE-----TTGVLKMWNL 2013  
Db 1706 ATTAMGSPKEVIF-----SDITENSATVSWRAPTAQVESFRITVYPITGGTPSMVT 1757  
Qy 2014 QSGGFSCTIRYKRGPLVDKQI-----YRFSE-----EGMVNARFDYTYH 2053  
Db 1758 D--GTKQTQRLVKLIPGEVLYLSIAKGFEESEPVSGSFTTALDGPGLVTANI----- 1810  
Qy 2054 DNSPRIASIKPVISETPLPVDLYRYDEISGKVEHFGKF---GVYIYDINQIITTAVMTLS 2110  
Db 1811 TDSEALARWQPAIA---TVDSYVISVTGEKVPEITRTVSGNTVEYALTDLPEATEVTL- 1865  
Qy 2111 KHPDTHGRIK-----EVQYEMFRSLMYMTVOYDSMGRVIKREL 2149  
Db 1866 RIFAEPQKSSTTAKFTTDLDSPLDLTATEVQSE---TALLTRWP----- 1909  
Qy 2150 KLGPYANTTKYTYDYG--DGOLQSVAVMNDRTWRSYDLNGLNHLNPGNSVRLMPLRYD 2208  
Db 1910 ---PRASVTGVLVYESVDGTVKEVIGPDTYSVLADLSPSTHY-----TAKIQALNGP 1961  
Qy 2209 LRDRI-----TRLGDVQYKIDDDGYLRCORGSDIFEYNSKGLLTRAYN----- 2250  
Db 1962 LRSMNIQTIETFTT--LLYPPFKDCSQAMLNGD----TTSGLTYTYLNGDKAQALEVFCDM 2016  
Qy 2251 --KASGMSVOVRYDGVGRR-----ASKYTNLGHHLQYFSDLHNPTRIITHVYHNSSEI 2302  
Db 2017 TSDGGGMIWFLRRKN--GRENFYQNNKAYAGFGDRREFFWJGLDNLNKRTIAQGOY----- 2070  
Qy 2303 TSLYDYLQGHFLFAMESSEYEEYVASDNTCTPLAVFSINGLMIR--QLQYATYAGEIYYDSN 2361  
2071 -ELRVDLRDH-----GETAFAPYDK-----FSVGDAKTRYKLKVEGYSGTAGDS- 2113  
Qy 2362 PDFQWVGFHGG 2373  
Db 2114 -----MAYHING 2119

RESULT 9  
JQ1322  
tenascin precursor - mouse  
N:Alternate names: contactin; hexabrachion  
C:Species: Mus musculus (house mouse)  
C:Date: 10-Sep-1999 #sequence.revision 10-Sep-1999 #text\_change 16-Jun-2000  
C:Accession: JQ1322; A37936; B37936; S14571; S50209  
R:Saga, Y.; Tsukamoto, T.; Jing, N.; Kusakabe, M.; Sakakura, T.  
Gene 104, 177-185, 1991  
A:Title: Murine tenascin: cDNA cloning, structure and temporal expression of isoforms.  
A:Reference number: JQ1322; MUID:92009211  
A:Accession: JQ1322  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-2019 <SAG>  
A:Cross-references: GB:D90343; NID:g220609; PIDN:BAA14355.1; PID:g220610  
A:Experimental source: cell line 2H6GR  
A:Note: the authors translated the codon ATG for residue 60 as Trp  
R:Weller, A.; Beck, S.; Ekblom, P.  
J. Cell Biol. 112, 355-362, 1991

A:Title: Amino acid sequence of mouse tenascin and differential expression of two ten  
A:Reference number: A37936; MUID:91107734  
A:Accession: A37936  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-201, 'E', 203-317, 'S', 319-620, 622-1010, 'N', 1012-1018, 'S', 1020-1024, 'H', 10  
A:Cross-references: GB:X56304  
A:Accession: B37936  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-201, 'E', 203-317, 'S', 319-620, 622-1010, 'N', 1012-1018, 'S', 1020-1024, 'H', 10  
A:Cross-references: GB:X56304  
R:Weller, A.; Beck, S.; Ekblom, P.  
submitted to the EMBL Data Library, August 1990  
A:Description: Amino acid sequence of mouse tenascin and differential expression of tw  
A:Reference number: S14571  
A:Accession: S14571  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-201, 'E', 203-317, 'S', 319-620, 622-1010, 'N', 1012-1018, 'S', 1307-2019  
A:Cross-references: EMBL:X56304; NID:g54768; PIDN:CAA39751.1; PID:g54769  
R:Glumoff, V.; Savontaus, M.; Vehanen, J.; Vuorio, E.  
Biochim. Biophys. Acta 1219, 613-622, 1994  
A:Title: Analysis of aggrecan and tenascin gene expression in mouse skeletal tissues  
A:Reference number: S50206; MUID:95035091  
A:Accession: S50209  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 46-146 <GLU>  
A:Cross-references: EMBL:X80281  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994  
C:Superfamily: tenascin; EGF homology; fibrinogen beta/gamma homology; fibronectin ty  
C:Keywords: alternative splicing; calcium binding; cell adhesion; duplication; extrac  
F:1-22/Domain: signal sequence #status predicted <SIG>  
F:23-2019/Product: tenascin, long splice form #status predicted <MAT>  
F:23-1071.1527-2019/Product: tenascin, short splice form #status predicted <MAT2>  
F:408-434/Domain: EGF homology <EGF>  
F:622-703/Domain: fibronectin type III repeat homology <FN3A>  
F:711-793/Domain: fibronectin type III repeat homology <FN3B>  
F:802-884/Domain: fibronectin type III repeat homology <FN3C>  
F:892-976/Domain: fibronectin type III repeat homology <FN3D>  
F:984-1064/Domain: fibronectin type III repeat homology <FN3E>  
F:1073-1155/Domain: fibronectin type III repeat homology <FN3F>  
F:1346-1428/Domain: fibronectin type III repeat homology <FN3G>  
F:1437-1519/Domain: fibronectin type III repeat homology <FN3H>  
F:1527-1608/Domain: fibronectin type III repeat homology <FN3I>  
F:1616-1696/Domain: fibronectin type III repeat homology <FN3J>  
F:1704-1784/Domain: fibronectin type III repeat homology <FN3K>  
F:1799-2007/Domain: fibrinogen beta/gamma homology <FBG>  
F:38,166,184,327,788,1018,1079,1093,1119,1184,1210,1275,1301,1354,1364,1394,1627,1878

Query Match 4.7%; Score 679.5; DB 1; Length 2019;  
Best Local Similarity 20.3%; Pred. No. 1.7e-27;  
Matches 421; Conservative 251; Mismatches 686; Indels 715; Gaps 98;  
Qy 535 CPRNCHNGECVSGVCHCFPGFLGADCAKACPVLCGSGNGQYSGKTCQCYSGWKGAEC-D 593  
Db 346 CPNDGCGQCEEGOCVNEGFACDCSEKRCPADCHRRGRLNGQCEDDGFTGADCGD 405  
Qy 594 VPMNQCIDPSCGGHSGCIDGNCVCSAGYKGBEHCBEVDCLDPTCSSHGVCVNGECLCSPGW 653  
Db 406 L---QCPN-GCSGHGRCVNGQCVDGEGTGBDCSQRRCPN-DCHNRGLCVQKGCICEQGF 460  
Qy 654 GGLNCELARVQCPDQCSGHGTYLPDTGLCSDPNWMDGCSVEVCSVDCGTHGVCIGAC 713  
Db 461 KGFDG---SEMSPNDCHQHGRCV---NGMCICDDDTGDECDRDRRCPRDCSQRGRVCDGQC 516  
Qy 714 RCEEGWTGAACDQRVCHPRCTEHCCTCKDGKCEGECNGEHCCTIGRQTAGTGTGCGCPDLC 773  
Db 517 ICEDGFTGPDCAELSCPSDCHGRCVNGQCICHEGFTGDKCEQR-----CFSDC 567  
Qy 774 NNGRCTLGQNSWQCVQOTGWRGPGCNVAMETSCADNKDNEGDLVDCLDPCCLQSAQC 833

Db 568 HGQGRCEG---QCICHEGFTGLDCG---QRSCPNDCSNOGO-----CVSGRC- 609  
QY 834 NSLLCRGRDPLDIIOGGQDMPAPKSFYDRIKLAGKDSHIIPGENPNSLSVLIRG 893  
Db 610 -----ICNEGVT----- 616  
QY 894 QVVTDTGLPLGVNVFVYKPKYGTITRQDGTFDLIANGASLTLHFRAPFMSQERTV 953  
Db 617 -----GIDCSEVSPK-----DLIV--TEVT-----EETV 639  
QY 954 WLPW-NSFVMTDLNKMTEENS-----IPCDLSGVRPDPPIIISPLSTFFSAAPG 1004  
Db 640 NLAWDNEMRVETYLIMYTHADGLEMQFRVPGDQSTTIR-----ELEPG 685  
QY 1005 QNPVETOVLEHEIELPGSNVRLVSSRTAGYKSLKLTITQSVPLNLRVHLMAV 1064  
Db 686 VEYFIRVFAILENKRSP-----VSARVAY-----LPAP 715  
QY 1065 EGHLFQKFSQASPNLASTFIWKTDAYQORVYGLSDAVVSGVEYETCPSLIWE----- 1119  
Db 716 EGLKPKSIKETSVEVE---WDPLD-----IAFE-----TWEIFERN 748  
1120 -----KRTALQGFELDPNSLGGHSLDKHHIL--NVKSGILHKT 1157  
Db 749 MNKEDEGITKSLRRPETSRYOTGLAPGQYEYS-----LHTVKNTRGPGKLYT 799  
QY 1158 G-----ENQFLTQQAITSIMNGRRRSISCPSCNGLAEGNKLLAPVALAVGI-- 1206  
Db 800 TTRLDAPSHIEVKDVTDTTALTWF-----KPLAE-----DSIELSGIKD 841  
QY 1207 -DGLSVGDFNFIIRIRFPNRNVSILE-----LRNKEKHSNNPAHKYL-AVDPPVSGSLY 1260  
Db 842 VPGDRTTIDLTHEDNOYSIGNLRPDPEYEVSLISRRVDMASNPAKETFTGLDAPNLR 901  
QY 1261 VSDTNS-----RRV-----SRVK--SLSGTKLAGNSEVAVGCEQCLPDEARCGG 1306  
Db 902 VSQDMSITLEWRNVKADISRYIKYAPISG---GDHAEIDVPKSOQA----- 946  
QY 1307 GRAIDATLMSPGIADVKNGLMVFVDATMIRKVDONGIISTLLGSNDLFAVRPLS-CDSS 1365  
Db 947 -----TTKTTLGL---RGTGYGCVSAV-KGDRESDPATTNAATEIDAPKDLRSETT 997  
QY 1366 MDVAQVRLEWPTDLAVNPMDSLYLVLENNVILRITENHQVSIAGRMHCQV----- 1417  
Db 998 QD---SLTFEWTPLA--KFD-----RYRLNY--SLPTGOSMEVQLPKDATSHV 1039  
QY 1418 -----PGIDYLSKLAIHSALESASAIASHGTGVLYITETDEKINLRQVTTNGEICLL 1472  
Db 1040 LTDLEPGQET-----VLLIAEKGRHKSPARVKASTEVPSL 1077  
1473 AGAASDCCKNDVNCYSGDDAYATDAI-----LNSPSSLAVAPDGTIYIA 1519  
Db 1078 ENLTVTEAGWDGLRLNWTADDLAYEFVLOVEANNVTAHNTVPGNLRRAADIPGLKVA 1137  
QY 1520 DLGNTRIRAVSK--NKPVLNAPNOVEAASPGQELVYFNADGHQVTVSLVTEYLYN-- 1575  
Db 1138 TSYRVSIYGVARGYRTPVLSA-ETSGTPTNLGEVTAEV-GWDALTLNMTAPEGAYKNF 1195  
QY 1576 FYTSTDNDVTELDNNGNSLKRIRDSGMPRHLLMPDNOIITLVCTNGGLKVVSTONLE 1635  
Db 1196 FTQVLEADTQTQVN-----LTV--PGGLRSVDLPGLK 1226  
QY 1636 LGLMTVDGNTGLLATKSDGTGWTTFYDDHEGRLTNVTRPTGVVTSLHREMEKSITIDIE 1695  
Db 1227 AA-----TRYI-----TLRGVTQDFGA-----PLSVEV- 1251  
QY 1696 NSNRDDVTVIINLSVSEASYT-----VVQDVNRNSVOLCNN-----GTL 1735  
Db 1252 ---LTEDPOLGGLSVTEVSWDGLTLNWTDDLAYKHFVVQVEANNVBAQNLTVPVPSL 1308  
QY 1736 RVNYANGMG-----ISPHSEPHVLAGTIT---PTIGRCNLSLPMENGLNSTEW 1780

Db 1309 RAVDIPGLKADTPYRVSIYGIQYRTPMLSTDVSTAREPEIGNLNVSDVTKPSFN-LSW 1367  
QY 1781 RLKKEQIKKVTIFGRKLRVHGRNLLSIDYDRNIRTEKIYDDHRKFTLRIIYDQVGRPEL 1840  
Db 1368 ---TATDGIFFMTIEI-----IDSNRLQTAENINISGAERTAH- 1404  
QY 1841 WLPSSGL-AAVNVSYFFNGLRAGLQGMASERTDIDKQRI---VSRMFADCKVMSY- 1894  
Db 1405 ---SGLPSPSTDFIVLVSGIAPSIKTIISTATTAEPEVDNLLVSDATPDGFRLSWTA 1460  
QY 1895 ---YLDKSMVLLQSQRQYIFEDSSDRLLAVTMPVSARHSMSTHTSICGIIRNIYPPES 1951  
Db 1461 DEGIIFDSFVIRIDRTKKQ-----SEPBISLPSPERTDITGLREATEYEIYGISRG 1514  
QY 1952 NASVIFYDSDGRILKTSFLGTGROVYKYGKLSKLSSEIYVDYSTAVTFG----- 2000  
Db 1515 RR-----SQPSATATTAMGSPKEIMF-----SDITENANVSWRAPTAQVESFRI 1560  
QY 2001 --YDETTGVLMKVNLOSQGSFCTIRYRKIGPLVDKQIYRES-----EE----- 2041  
Db 1561 TVVPMTGGAPSMVTVD--CTDTETRLVKLTGPGVE---YRVSVIAMKGFESDPVSGTLIT 1615  
QY 2042 ---GWNARFDYTHHNSFRIASIKPVISETPLPVDLY-----RYDEISGKVEHF 2088  
Db 1616 ALDGPGLLIAMI-----TDSEALAMWQPAIA-----TVDSYVISYTGERVPEVTRTVSG- 1665  
QY 2089 GKFGVIYDINOIITFAVMTLSKHFDTHGRK-----EVOYEMF 2127  
Db 1666 ---NIVEYELHLEPATEYELS-IPAEGQOKSSITATKFTTDLDSPREFTATEVQSE-- 1719  
QY 2128 RSLMVMVTVOYDSMGKRVIKRELKLPYANTKYTYDYDG-DGOLQSVAVNDRPTWRYSYD 2186  
Db 1720 TALLTWRP-----PRASVTGYLLVYESVDGTVEKVIYVGPDTTSYSLAD 1762  
QY 2187 LAGNLHLLNPGNSVRLMLRYDLRORI-----TRLGDVOYKIDDDGYLCORG-----S 2234  
Db 1763 LSPSTHY-----SARIQALSGSLRSKLITQITFTIG-LLYPPFRD---CSQAMLNGDTS 1813  
QY 2235 DIFEVNSKGLLTRY-----NKASGWSVQVRYDCVGRR-----ASYKTLGLHHLQYF 2281  
Db 1814 GLYTIYINGDKTOALEVYCDMTSDGGWIVFLRRKN-GREDFYRNWKAAGGDRREF 1872  
QY 2282 YSDLINPTRITHYVNHNSSEITSLYDYGLOGLHFAMESSEGEYEVASDNTGTPLAVFSIN 2341  
Db 1873 WGLGDLNLSKITAQQGV-----ELRVDLQDH-----GESAYAVYDR-----FSVG 1911  
QY 2342 GLMIK-OLQYATAYGEIYYDSNPDFQWVIGFHGS 2373  
Db 1912 DAKSRTKLKVEGYSGTAGDS-----MNYHNG 1937  
RESULT 10  
S19694  
tenascin precursor - pig  
N:Alternate names: contactin; hexabrachion  
C:Species: Sus scrofa domestica (domestic pig)  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: S19694  
R:Nishi, T.; Weinstein, J.; Gillespie, W.M.; Paulson, J.C.  
Eur. J. Biochem. 202, 643-648, 1991  
A:Title: Complete primary structure of porcine tenascin. Detection of tenascin transcr  
A:Reference number: S19694; MUID:92104189  
A:Accession: S19694  
A:Molecule type: mRNA  
A:Residues: 1-1746 <NIS>  
A:Cross-references: EMBL:X61599; NID:92124; PIDN:CAA43796.1; PID:92125  
C:Superfamily: tenascin; EGF homology; fibrinogen beta/gamma homology; fibronectin ty  
C:Keywords: alternative splicing; calcium binding; cell adhesion; duplication; extrac  
F:1-22/Domain: signal sequence #status predicted <SIG>  
F:23-1746/Product: tenascin #status predicted <NAT>  
F:346-372/Domain: EGF homology <EGF>  
F:377-403/Domain: EGF homology <EGF2>  
F:622-703/Domain: fibronectin type III repeat homology <FN3A>

F:701-793/Domain: fibronectin type III repeat homology <FN3B>  
F:802-984/Domain: fibronectin type III repeat homology <FN3C>  
F:882-976/Domain: fibronectin type III repeat homology <FN3D>  
F:984-1064/Domain: fibronectin type III repeat homology <FN3E>  
F:1073-1055/Domain: fibronectin type III repeat homology <FN3I>  
F:1154-1245/Domain: fibronectin type III repeat homology <FN3J>  
F:1254-1336/Domain: fibronectin type III repeat homology <FN3K>  
F:1343-1423/Domain: fibronectin type III repeat homology <FN3L>  
F:1431-1511/Domain: fibronectin type III repeat homology <FN3M>  
F:1526-1734/Domain: fibronogen beta/gamma homology <FBG>  
F:38,166,184,327,788,1034,1079,1121,1354/Binding site: carbohydrate binding site

Query Match 4.5%; Score 656; DB 1; Length 1746;  
Best Local Similarity 21.0%; Pred. No. 2.4e-26;  
Matches 379; Conservative 211; Mismatches 642; Indels 57

535	CPNRCHNGCEVSGVGHCHFPGFGLGADCAKAAACPVLCGSGNQYYSKGTCCQCYSGWKGAECVD	594
315	CPKDCDFGRGRIINCTYCDGFGFEGEDCGLRCLPHGCRGRGRCBEGQCVCEGDEGAFGADCS-	373
595	PMNOCDIPS--CGHGSGCIDGNCVCSAGYKGEHCHEEVDCLDPTCCSSHGVCNVECLCSQMW	653
374	-ERRC--PSCDHNRCRLDGRCEBDDGDFEBDCGEURC--PGGCSHGRCVNCQVCDEGR	429
654	GGNCELARVQCQDQCSGHHGYLPDPTGLCSDPNMNGPDCSEVCSVDCTGHVCIIGGAC	713
430	TGEDC--SOLRCPNDCHGRGRCV--QGRCEBHGFGYDCSEMSCPHDCHQHGRCVNGMC	485
714	RCEEGWTGAACDORVCHPRCIENGHTCKDGKCEKCEBNGWGEHCTIGRTAGTETDGGPDLC	773
486	VCDGYTGEDCRELRCPGDCSQBRGCVDRGRCVEHGFAGPDCA-----DLACPSC	536
774	NGNRCTLQGNQWCQVCOTGWRGPGCNVAMETSCADKNONEGDLVDCLODPC-----	826
537	HGRGRCVNG---QCVCHEGFTGKDCG---QRRCPG--DCHGOG--RCVDGQCVCHEGFT	585
827	---CLQSACONSLLCRGSRDPLDIIQOGQTDWPAKVSFYDRIKLLAGKOSTHIIPGENPF	883
586	GLDCGQRSCFN-----DCSNWGO-----	603
884	NSSLVSLIRGOVVTDTGTPLVGVNVSFVKYKYGTITRQDGTDFDLIANGGASLTUHFER	943
604	-----CVSGRCICNEG--YSGEDCSQVSPK-----DLIV--TEXT-----	635
944	APFWSQERTVNLPW-NSFYAMDTLVWKTEBNS-----IPCDLSGFGVRPDPPIIISPP	994
636	-----EETVNLAWDNEMRVTEYLIVVTPTHEDGLENQFRVPGDQSTTIR-----	680
995	LSTFFSAACQNPITVPETOVLHHEEIELPGSNVLYRLSSRTAGYKSLIKITWTQSTVPLN	1051
681	-----ELEPGVEFIRVFAILENKKSIIP-----VSARVAY-----	711
1055	LIRVHLMAVVEGHLFOKSFQASPNLASTFTWKTDTAYGORYVGLSDAVVSQFEYETCPS	1111
712	-----LPTPEGLKFKSIKETSVEVE---WDPLD-----IAFE-----	740
1115	LILWE-----KRTALQGFELDPNSLGGWSLOKHHLNWK	1141
741	--TWEIFIRNNKDEGEITKSLRRPETYRTQGLAPGOEYEIS-----LHI--VK	787
1150	SGILHKGTGENOFLTQPAITISIMGRRRSICSPS---CNCLAEKNLLA---PVALA	1201
788	NNT--RCPGLKRVTTIR-----LDAPSQTEAKDVDTTALITWFKPLAEI	830
1204	VGDGSLYVGDFNYIRRIFFPSNRVNTSILELRNKEFKHSNNPAHYLAYDVPVSGSLYSD	1261
831	DGIELTYGKD-----VPGDRTT--IDLTHEENQYS-----IGNL-KPD	866
1264	TNSRRIRVKSLSGTKDLACNSEWVAGTGBOCLPFDEARCGCGKKAIDATILMSPRGIAVD	1321
867	TE---YEVSLISRRADMSN-----PAKE-----FTFTGLDAPRN-----	898

RESULT 11

RESOLUT  
T09070

probable tenascin X - mouse

C; Species: Mus musculus (house mouse)

0  
2  
2  
0  
=

0  
2  
3  
0  
(

2  
3  
4  
5  
0  
2  
=

0  
3  
2

.  
2  
0  
4  
0  
2  
4

7  
6  
)

C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jan-2000  
C:Accession: T09070  
R:Rowen, L.; Mahairas, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Lasky, S.; Loretz, C.; S. submitted to the EMBL Data Library, October 1997  
A:Description: Sequence of the mouse major histocompatibility locus class III region.  
A:Reference number: 216543  
A:Accession: T09070  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-4006 <ROW>  
A:Cross-references: EMBL:AF030001; NID:g2564945; PID:g2564958  
C:Genetics: TNX  
A:Gene: TNX  
A:Map position: 17  
A:Introns: 124/1; 735/1; 773/3; 826/1; 914/1; 1037/1; 1135/1; 1232/1; 1329/1; 1440/1; 1501/1; 3115/1; 3208/1; 3302/1; 3405/1; 3517/1; 3558/1; 3606/1; 3646/1; 3694/1; 3737/3; 3737/3; 3737/3  
C:Superfamily: tenascin-X; EGF homology; fibrinogen beta/gamma homology; fibronectin type C; keywords: extracellular matrix  
F:422-448/Domain: EGF homology <EGF>  
F:826-906/Domain: fibronectin type III repeat homology <3FR>  
F:3789-3997/Domain: fibrinogen beta/gamma homology <FBG>

Query Match 4.3%; Score 632.5; DB 2; Length 4006;  
Best Local Similarity 24.0%; Pred. No. 1.7e-24;  
Matches 253; Conservative 108; Mismatches 358; Indels 337; Gaps 49;

QY 531 SYQDCPRNCHGNGCVSGVCHYCPGFLGADCAKACPVLCGNGQVSKGTCOCYSGWKGA 590  
DB 418 STRACPRDCRGRCRGEDGVCVCHAGYSGEDCGVRCPGDCRGNCESGRVCWFGYGR 477  
QY 591 ECDVPMNQIDPSCGGHSCIDGNCVCSAGYKGEHEEYDCLDPTCSSHGVCNGECLCS 650  
DB 478 DCG---TRACPGDCRGRCVDCGVCNPGFTGEDCGSRRC-PGDCRGHGHGCVCA 533  
QY 651 PGWGLNCELARVQCDQCSHGTYLPTDGLCSDPNWMGPDSCVSVDCGTHGVCIG 710  
DB 534 VGYSGDDCSTR--SCPSDCRGRCQL--NLGCECEGYSGEDCGIRRCPRDCSQHGVCD 589  
QY 711 GACRCEEGWTGAACQQRVCHPRCIIEHGTCKGCKCEGNGEHCITIGRTAGTDTGCP 770  
DB 590 GLCMCHAGYAGEDCSRTCPADCRRCRGCEDCRCVNGYGPACA-----TRTCP 640  
QY 771 DLGNGNGRCTLGQ-----NSQCVCQTCWRGPG 798  
DB 641 ADCRGRCVQVCMCYGVYSGEDCGQEPASACPGCGPRELCRAGQCVCVEGFRGP 700  
QY 799 CNVAMETSCADKNEGDLVCLDPCCLQACQNSLLCR-----GSRDP----- 844  
DB 701 C--AIQT-CPG-----DCSRGECIQRC-----VCQEGYAGDDCGEIPAIQNM 743  
QY 845 LDIIQGG--QTDW-----PAYKSFYDRKLLAGKD- 872  
DB 744 MHLLETTVTEWTRAPGPDVAYEIQIPMTSGVSPPTTARVPSSASAYDQRLAPGQDY 803  
QY 873 ---STHIIPEEN---PFNSLSVLIRG---QVVTDTGT-----PLGVNVSFVKPK 915  
DB 804 QVTVALRGTSPPASKTITTMIDGPDLRVAVYPTTLDLSWLRPOAEVDVRFVYSVS 863  
QY 916 YGYTITR-----QDGT--FDLIANGASLTUHFERAPMSOERTVWLPWNSFYAMDITLV 967  
DB 864 AGNQRVLEVPPEADRTQLDMPGVEYVVTVAERGHVASYPASI----- 909  
QY 968 MKTEENSPCDLSGFVRPDPPIISPLSTFFSAAPGNPIVPETQVLHEETELPGSNVK 1027  
DB 910 -RANTGSLPSGLLEATDEPPP--SGPSTQGAQAP---ILILEHPLGE----- 952  
QY 1028 LRYLSRAGYKSLKIMTQSTVPLNLRVHLMVA-----VEGHLFOKSTQASPN 1078  
DB 953 LKVLGRDKAG---RLSVAMTAQPDPSFAHQLRMQVAEGFPWAHEELLPGDVQQAALVPPPPP 1009  
QY 1079 LASTFIWDKTDAYGQVYGLS--DAVVSVGFVEYTC-----PSLILWEKRTALLQG 1127

DB 1010 GA-----PYKULFLHGIITPGGKISVPIVYQGIIMRAQEQPGKPSV---OPRLG----- 1053  
QY 1128 FELDPNSLGGSLDKHHLNLSKSGILHKGTEGNO---FLTQO-----PAITTSINGNRR 1179  
DB 1054 -ELVTGTLTSDSL-----LLHWTVPGEFDSFLIQYKDKGQPAI-SVEGPPQRS 1100  
QY 1180 RSISCPSCNGLAEGNK-----LLAPVALAVGID-----GSLYVGD 1214  
DB 1101 TPIS-----GLEPGRKYFILYGLIGKKRHGPLMAEAKILIQSDPDGSPRRLGELWTD 1155  
QY 1215 -----FNVIIRIFPSRNVTSILELNKEFKHSNNPAHKYYLAVDPVSG---SLYSDTNS 1266  
DB 1156 PTPHSLHLSWTVLGQGFDSFVQYRDKE-----GQP-----RWVPEGPDRSWISPLDP 1205  
QY 1267 RRIYRVKLSGTLKOLAGNSEVAVGTGEQCLPDEARCGDGGKAIDATLMSPRGIADVKNK 1326  
DB 1206 NRKYRF-----TLF---GIANKRY 1222  
QY 1327 LMYFVDATMIRKVDQNGIITLLGSNDLTAVRPLSCDSMDVAQVRLPMDTDLAVNPMDN 1386  
DB 1223 GPLTADGTTAAETKESESSEPPRLGELTVGTVP---DS-----URLSW--TVARGPFD 1271  
QY 1387 SLYVLNNVILRITENHOVSI-IAGRPMHCQVPGID 1421  
DB 1272 FV-----ILYKDAQGQPSVPIEGDENEVTVPGLE 1301

RESULT 12  
T12457  
hypotheical protein DKFP56400423.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C>Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 23-Jul-1999  
C:Accession: T12457  
R:Wambutt, R.; Heubner, D.; Meves, H.W.; Gassenhuber, J.; Wiemann, S. submitted to the Protein Sequence Database, June 1999  
A:Reference number: 217524  
A:Accession: T12457  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-184 <WAM>  
A:Cross-references: EMBL:AL080120  
A:Experimental source: fetal brain; clone DKFP56400423  
C:Genetics:  
A:Note: DKFP56400423.1

Query Match 4.3%; Score 627.5; DB 2; Length 184;  
Best Local Similarity 62.4%; Pred. No. 1.8e-26;  
Matches 116; Conservative 39; Mismatches 28; Indels 3; Gaps 1;

QY 2548 ASEDSRKVASVLNNAYLDKMHYSIEGKDTYFVKIGSADGLVTLGTTIGRKVLESVN 2607  
DB 2 ANEDGRRVAATLDHAYLENLHFTIDGVDTHYFVKPGPSEGDLATLGLSGRRRTLENGV 61  
QY 2608 VTVSPTLLVNGRTRFRFTNIEFQYSTLLSIRYGLTPTDDEEKARVLDQARQALGTAW 2667  
DB 62 VTVSINTVLNGRTRYTDIQLQYGCALCNTRYG---TTLDEEKARVLELARQAVROAW 118  
QY 2668 AREQOKARDREGSRRLWTEGKQQLLSTGRVQGYGVYVLPVQYPELADSSSNITQFLRQ 2727  
DB 119 AREQORLEBEGELRAWTEGKQQLVSTGRVQGYGVYVLPVQYPELADSSSNITQFLRQ 178  
QY 2728 NEMGKR 2733  
DB 179 SEMGRR 184

RESULT 13  
A40701  
tenascin-X precursor - human  
C:Species: Homo sapiens (man)  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Dec-1999  
C:Accession: A40701; A33725; C42175



Db 1411 APEPHGELTVEATGHTLH-LSMVTEGEFEISFIQYTRDQQLQMVIRIGGRNDITLS 1469  
QY 1489 CYSDDAYATDALNSPSSILAVAPDCTIYIADLGNIRIRAVSKN-----KPVLNAP 1539  
Db 1470 GLESDHRYLV-----TLYGFSOG-----KHVGPVHVHVALTEPPATPEPIKPRUGEL 1517  
QY 1540 NOYEAAAPGEQEELYVFNADGIIHQYTVSLVTGEVLYNFTYSTDNVDPTELIDNNGNSLKIR- 1598  
Db 1518 TVTID-ATPDSLSL-----SWTVP--EGGF-----DHFLVQYRNGDGQPKAVRG 1557  
QY 1599 -RDSSOMPRLHMPDNQIITLTVGTNGG-----LKVSTQNLE-----1635  
Db 1558 PGHEGVTTISGLEPDHKYKMNLYGFHGGQRMGPVSVYTAABEETPSPTEPSMEAPEA 1617  
QY 1636 ---LGLMYDGNGLATKSDETGWT-----TFYDYDHEGRLTNVTRPTG-----1677  
Db 1618 EPLLGELITVTS-----SPDSLSLSTVTPQGREDSTVQYKORDGR-PQVVRVGGPEERS 1672  
QY 1678 -----VVTSLHREMEKSIITDIENSRRDDVTITNLSSVEASYTVVQDVQRNSYQLCNN 1732  
Db 1673 PDAPLAKRLGOMTVRDITSD-----SLSLSWTVPQGFQDH-----1708  
QY 1733 GILRVNYANGMISFSEPHVL-----AGTITPTIGRCNISLPMENGLNSIEHRLK 1784  
Db 1709 ---FLVQFKNGDG-----QPKALRVPHGEDGABEETPS-----PTEPSMEAPE--PPE 1751  
QY 1785 EQIKGKVTIFG 1795  
Db 1752 EPLLGELITVGT 1762

## RESULT 14

A43902  
tenascin - eastern newt (fragment)  
C:Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 12-Feb-1999  
C:Accession: A43902  
R:Onda, H.; Poulin, M.L.; Tassava, R.A.; Chiu, I.M.  
Dev. Biol. 148, 219-232, 1991  
A:Title: Characterization of a new tenascin cDNA and localization of tenascin mRNA during gastrulation in the embryo of the newt, *Notophthalmus viridescens*  
A:Reference number: A43902; MUID:92038434  
A:Accession: A43902  
A:Molecule type: mRNA  
A:Residues: 1-647 <OND>  
A:Cross-references: GB:M76615  
A:Note: sequence extracted from NCBI backbone (NCBI:64543, NCBI:64547)  
C:Superfamily: tenascin; EGF homology; fibronogen beta/gamma homology; fibronectin type I  
C:Keywords: extracellular matrix; glycoprotein; tandem repeat  
F:287-313/Domain: EGF homology <EGF>  
F:316-427/Domain: fibronectin type III repeat homology <FN1>  
F:515-517/Domain: fibronectin type III repeat homology <FN2>

Query Match 4.1%; Score 604; DB 2; Length 647;  
Best Local Similarity 28.2%; Pred. No. 2.7e-24;  
Matches 153; Conservative 75; Mismatches 207; Indels 108; Gaps 23;  
QY 535 CPNCHNGECVSGVCHCPFLGADCAKAAACPVLCSGNGQYSGKTCQCYSGWKGAECV 594  
Db 70 CPNDCFDRCRCINGVCFDEGTGEGDGCGLTCTPNCNNRGRVNGLCVCDGFGQDDCS- 128  
QY 595 PMNQCTDPSCGGHSGCIDNGCVCSAGYKGEHCEVDCLDPTCSSHGVGVNGECLSPGNG 654  
Db 129 -ELRCFN-DCNDRGRVNGKVCKEGFMGEDCADLRCPN-DCNNRGRVNGQVCVDEGFM 185  
QY 655 GUNCELARVQCDCSGHCTYLPDTGLCSLCPNPMGPDSCSVEVCSVDCGTHGVCTGGACR 714  
Db 186 GEDC--SDLRCPGDCNNRGRV--NQQVCDEGFGEDGELRCPPDCNNRGRVNGQCI 241  
QY 715 CEEGWTGAACDORVCHPRICIEHGTCKDGKCEGNGEHCITIGQTACTETDGCPLCN 774  
Db 242 CDEGFMGNGELRCPNDCNKRGRVNGQICDDGFKGEDCSLR-----CPDCN 292

QY 775 GNGRCTLGQNSQCYCQTCQGWGPGCNVAMETSCADNKNEDGLVDCLDPDCL--QSAC 832  
Db 293 DRGRCLNG--QCVAEGFTGEND--SLACLN-----CNDRGCLVNGQCV 335  
QY 833 QNSLL--CRGRDPLDIIIOGQDMPAVKSYDRIKLL-----AKDSTHIIP 878  
Db 336 EGFGLGEDCSEVSPKDLTVDTTQSVNLEWANEKMTVEYLITYPTSPGGLDLDVRP 395  
QY 879 GENPNSLSVLIRQVVTDTGTPLVGVNVSEVKPKYCYTITRODGTDLTANGASL- 937  
Db 396 GDOT-TATIQLEPG-----VEYFVRVFAIRNQRSIPVSARVATHLP 437  
QY 938 ---TLHFERAPMSOERTVWLPWN---SFYAMDILVMKT-EENSIPSCDLSGFRVPDPI 989  
Db 438 TTDDLRFKSV---KETSVEVEMDPLDISFDTWLIIINTKEENGEISLSQ--RP---487  
QY 990 IYSSPLSTFFSAAPQNPVPTQVLHBEIELPGSNVKLYLSSRTAGYKSLKLTMTQS 1049  
Db 488 -VTSYVQT--GLAPGET-----YNFSIHVVKNSTRG-PGLAKVTITRL 526  
QY 1050 TVP 1052  
Db 527 DAP 529  
RESULT 15  
T42629  
tenascin-X - bovine  
N:Alternate names: flexilin  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 21-Jul-2000  
C:Accession: T42629  
R:Eleftheriou, F.; Exposito, J.Y.; Garrone, R.; Lethias, C.  
J. Biol. Chem. 272, 22866-22874, 1997  
A:Title: Characterization of the bovine tenascin-X.  
A:Reference number: Z22180; MUID:97426436  
A:Accession: T42629  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-4135 <ELE>  
A:Cross-references: EMBL:Y11915; NID:g2462978; PIDN:CAA72671.1; PID:g2462979  
C:Gene: TN-X  
C:Genetics:  
C:Superfamily: tenascin-X; EGF homology; fibronogen beta/gamma homology; fibronectin type I  
C:Keywords: extracellular matrix; glycoprotein; heptad repeat  
Query Match 4.0%; Score 590; DB 2; Length 4135;  
Best Local Similarity 26.4%; Pred. No. 3.4e-22;  
Matches 172; Conservative 63; Mismatches 216; Indels 200; Gaps 27;  
QY 532 VQDCPRNCHNGECVSGVCHCPFLGADCAKAAACPVLCSGNGQYSGKTCQCYSGWKGAECV 591  
Db 400 VRSCPSDCNQRCEGRCVCHPFGYSGPDGACAPRDCRGRGRCNGVGVCHAGYSGED 459  
QY 592 CDV-----PMNQCI-----DPSCGGHSGCIDNGCVCSAGYK 623  
Db 460 CGVRSCPGDCRRRRGRRCESGRVCVPGYTGRCGTRACPGDCRGRVDCRVCNPGFAG 519  
QY 624 EHCEVDCVLDPTCSSHGVGVNGECLSPGNGGLNCELARVQCDCSGHGTYLPDTGLCS 683  
Db 520 EDCGSRRC-PGDCRGRGRCGDCVCDVGYEGEDC--GKRSCPRCQGRGQCL--EGRCV 574  
QY 684 CPNMMGPDSCSVEVCSVDCGTHGVCTGGACRCEEGWTGAACDQVCHPRICIEHGTCKDGK 743  
Db 575 CDDGYEGEDCGVRRCPRCNQRGVCDGVCTCWEFAGEDCGLRVCPNCHRRGRCENGR 634  
QY 744 CECREGWNGEHCITIGQTACTETDGCPLCNNGNRCITLGQNSWQCVQCTGWRGPCN---800  
Db 635 CVCDSGYTGPSCA-----TRTCPADCRGRGRCVQG---VCVCHGVSGEDCGQEE 681  
QY 801 ---VAMETSCADNKNEDGLVDCLD-----PDCLQSAQNSLLCRG---SRDPLDIQQ 850

```
Db 682 PPASACPGCGP-RELCSAGCVCVEGFRGPDCAIQT-CFGD--CRGRGECREGSCVCQD 737
QY 851 G-----QTDW-----855
Db 738 GYAGEDCGEVPAGEGRMHLETTVTWTRAPGNVDAYEIQFIPTTEGASPPFTARV 797
QY 856 PAVKSFYDRIKLAGKD---STHIPGEN---PNSSLVSLIRG-----QVTTDGT---- 901
Db 798 PSSASAYDQRLAPGQEVQVTVRALRCTNNGPPASKTITTMIDGPQDLRVAVTPTTLEL 857
QY 902 -----PLGVNVVSFVKYKGYTITR-----QDGTG--DLIANGCASLTLHFERAPFMSQ 949
Db 858 NWLRPQAEVDREVVYSVSAGNQVRLEVSEADGTLTGLMPGVVEYVTVTAERGRAVS 917
QY 950 ERTVWLPWNFSFYAMDTLVMKTEENSIPSCDLSGFVRPDPIISSPLSTFTTSAAP--GONP 1007
918 PASIRA--NTGSSLSGLLGATDE-----PPP---SGPSTTQGAQAPVLOQRP 959
1008 -----IVPETQVVLHEEIE LPG 1023
Db 960 QELAEALRVLGKDKTGLRLRVAVTAQPDFTTHFQLRLRVPEGPGAHEEL-LPG 1009
```

Search completed: September 18, 2002, 11:09:55  
Job time: 265 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 11:08:25 ; Search time 23.68 Seconds  
(without alignments)  
4468.772 Million cell updates/sec

Title: US-09-800-198-8

Perfect score: 14581

Sequence: 1 MDVKDRHRSLTRGCRKREC.....ELADSSNQIFLRQNMGRK 2733

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

ched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	732.5	5.0	1808	1 TENA_CHICK	P10039 gallus gall
2	680	4.7	2201	1 TENA_HUMAN	P24821 homo sapien
3	656	4.5	1746	1 TENA_PIG	Q29116 sus scrofa
4	616	4.2	4289	1 TENX_HUMAN	P22105 homo sapien
5	463.5	3.2	2531	1 NTC1_MOUSE	Q01705 mus musculus
6	462.5	3.2	2531	1 NTC1_RAT	Q07008 rattus norv
7	450.5	3.1	2318	1 NTC3_MOUSE	Q61982 mus musculus
8	448.5	3.1	2444	1 NTC1_HUMAN	P46531 homo sapien
9	437.5	3.0	1064	1 FBPI_STRPU	P10079 strongyloce
10	428.5	2.9	2524	1 NOTC_XENLA	P21783 xenopus lae
11	417	2.9	2139	1 CRB_DROME	P10040 drosophila
12	417	2.9	2437	1 NOTC_BRARE	P46530 brachydanio
13	415.5	2.8	830	1 SREC_HUMAN	Q14162 homo sapien
14	413	2.8	2703	1 NOTC_DROME	P07207 drosophila
15	399	2.7	1295	1 GLP1_CAEEL	P13508 caenorhabdi
16	396.5	2.7	1376	1 CRBH_HUMAN	P82279 homo sapien
17	395.5	2.7	1408	1 SERR_DROME	P18168 drosophila
18	395	2.7	833	1 DL_DROME	P10041 drosophila
19	391	2.7	1964	1 NTC4_MOUSE	P31695 mus musculus
20	368.5	2.5	473	1 FP2_MYTGA	Q25464 mytilus gal
21	363	2.5	686	1 DLL4_MOUSE	Q31171 mus musculus
22	360.5	2.5	1429	1 LI12_CAEEL	P14585 caenorhabdi
23	350	2.4	570	1 FBP3_STRPU	P49013 strongyloce
24	346.5	2.4	714	1 WIF1_RAT	P97677 rattus norv
25	345	2.4	379	1 WIF1_MOUSE	Q89411 mus musculus
26	344	2.4	379	1 WIF1_HUMAN	Q39545 homo sapien
27	343	2.4	723	1 DLL1_HUMAN	Q00548 homo sapien
28	338.5	2.3	685	1 DLL4_HUMAN	Q9nr61 homo sapien
29	335.5	2.3	3075	1 LMA1_HUMAN	P25391 homo sapien
30	332	2.3	722	1 DLL1_MOUSE	Q61483 mus musculus
31	328.5	2.3	4660	1 LRP2_RAT	P98158 rattus norv
32	327	2.2	3672	1 LML2_CAEEL	Q21313 caenorhabdi
33	322.5	2.2	2334	1 WAPA_BACSU	Q07833 bacillus su

#### ALIGNMENTS

RESULT 1

ID	TENA_CHICK	STANDARD	PRT	1808 AA.
AC	P10039: P13132; 073584; 073585;			
DT	01-MAR-1989 (Rel. 10, Created)			
DT	01-MAR-1992 (Rel. 21, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Tenascin precursor (TN) (Hexabrachion) (Cytotactin) (Neuronectin) (GWMEM) (JI) (Mitochondinous antigen) (Glioma-associated-extracellular matrix antigen) (GP 150-225).			
DE	Gallus gallus (Chicken).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.			
OC	NCBI_TaxID=9031;			
OX	[1]			
RN	SEQUENCE FROM N.A.			
RP	TISSUE=Embryo;			
RC	MEDLINE=90030407; PubMed=2478295;			
RX	Spring J., Beck K., Chiquet-Ehrismann R.;			
RA	"Two contrary functions of tenascin: dissection of the active sites by recombinant tenascin fragments.";			
RT	Cell 59:325-334(1989).			
RL	[2]			
RN	SEQUENCE OF 27-722 FROM N.A., AND SEQUENCE OF 79-96.			
RP	TISSUE=Fibroblast;			
RC	MEDLINE=89030589; PubMed=2460335;			
RX	Pearson C.A., Pearson D., Shibahara S., Hofsteenge J., Chiquet-Ehrismann R.;			
RA	"Tenascin: cDNA cloning and induction by TGF-beta.";			
RT	EMBO J. 7:2977-2982(1988).			
RL	[3]			
RN	SEQUENCE OF 464-1018 AND 1412-1661 FROM N.A., AND SEQUENCE OF 852-868.			
RP	TISSUE=Embryo;			
RC	MEDLINE=88176910; PubMed=2451243;			
RX	Jones F.S., Burgoon M.P., Hoffman S., Crossin K.L., Cunningham B.A., Edelman G.M.;			
RA	"A cDNA clone for cytotactin contains sequences similar to epidermal growth factor-like repeats and segments of fibronectin and fibronogen.";			
RT	Proc. Natl. Acad. Sci. U.S.A. 85:2186-2190(1988).			
RL	CC - - FUNCTION: SAM (SUBSTRATE-ADHESION MOLECULE) THAT APPEARS TO INHIBIT CELL MIGRATION. MAY PLAY A ROLE IN SUPPORTING THE GROWTH OF EPITHELIAL TUMORS. IS A LIGAND FOR INTEGRINS ALPHA-8/BETA-1, ALPHA-9/BETA-1, ALPHA-V/BETA-3 AND ALPHA-V/BETA-6.			
CC	CC - - SUBUNIT: HEXAMERIC. A HOMOTRIMER MAY BE FORMED IN THE TRIPLE COILED-COIL REGION AND MAY BE STABILIZED BY DISULFIDE RINGS AT BOTH ENDS. TWO OF SUCH HALF-HEXABRACHIONS MAY BE DISULFIDE LINKED WITHIN THE CENTRAL GLOBULE.			
CC	CC - - SUBCELLULAR LOCATION: Extracellular matrix.			
CC	CC - - ALTERNATIVE PRODUCTS: 3 ISOFORMS; 230 KDA (SHOWN HERE), 200 KDA AND 190 KDA; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY ARE PRODUCED IN A TISSUE-AND TIME-SPECIFIC MANNER DURING DEVELOPMENT.			
CC	CC - - INDUCTION: BY TGF-BETA.			
CC	CC - - SIMILARITY: CONTAINS 13.5 EGF-LIKE DOMAINS.			
CC	CC - - SIMILARITY: CONTAINS 11 FIBRONECTIN TYPE III-LIKE DOMAINS.			

088516 mus musculus  
P98164 homo sapien  
O88671 rattus norv  
Q9ny77 homo sapien  
Q07954 homo sapien  
P24043 homo sapien  
Q00918 rattus norv  
Q00174 drosophila  
P19137 mus musculus  
Q61554 mus musculus  
Q61001 mus musculus  
P35555 homo sapien

34 312 2.1 592 1 DLL3\_MOUSE  
35 312 2.1 4655 1 LRP2\_HUMAN  
36 307.5 2.1 589 1 DLL3\_RAT  
37 302.5 2.1 618 1 DLL3\_HUMAN  
38 297 2.0 4544 1 LRP1\_HUMAN  
39 296 2.0 3110 1 LMA2\_HUMAN  
40 295.5 2.0 1712 1 LGFB\_RAT  
41 294.5 2.0 3712 1 LMA\_DROME  
42 291 2.0 3084 1 LMA1\_MOUSE  
43 288.5 2.0 2871 1 FBNI\_MOUSE  
44 287.5 2.0 3718 1 LMA5\_MOUSE  
45 284.5 2.0 2871 1 FBNI\_HUMAN

CC CC -1- SIMILARITY: CONTAINS A DOMAIN RELATED TO THE C-TERMINAL PART OF  
CC THE BETA AND GAMMA CHAINS OF FIBRINOGEN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M23121; AAA49086.1; -  
DR EMBL; X08031; CAB40811.1; -  
DR EMBL; X08030; CAA30824.1; ALT\_TERM.  
DR EMBL; J03641; AAA48748.1; ALT\_SEQ.  
DR EMBL; M20816; AAA48749.1; ALT\_SEQ.  
DR PIR; A30903; A30903.  
DR PIR; A31930; A31930.  
DR PIR; A33379; A33379.  
DR PIR; B33379; B33379.  
DR PIR; C33379; C33379.  
DR PIR; S01292; S01292.  
HSSP; P24821; 1TEN.  
InterPro; IPR000561; EGF-like.  
DR InterPro; IPR003961; FN\_III.  
DR InterPro; IPR002181; Fibrinogen\_C.  
Pfam; PF00008; EGF; 13.  
DR Pfam; PF00147; fibrinogen\_C; 1.  
DR Pfam; PF00041; fn3; 11.  
DR SMART; SM00181; EGF; 10.  
DR SMART; SM00001; EGF-like; 1.  
DR SMART; SM00186; FBG; 1.  
DR SMART; SM00060; FN3; 10.  
DR PROSITE; PS00022; EGF\_1; 14.  
DR PROSITE; PS01186; EGF\_2; 14.  
KW Glycoprotein; Cell adhesion; Repeat; EGF-like domain; Coiled coil;  
KW Extracellular matrix; Signal; Alternative splicing.  
FT SIGNAL 1 22  
FT PROPEP 23 33  
FT CHAIN 34 1808  
FT DOMAIN 118 142  
FT DOMAIN 176 188  
FT DOMAIN 188 219  
FT DOMAIN 219 250  
FT DOMAIN 250 281  
FT DOMAIN 281 312  
FT DOMAIN 312 343  
FT DOMAIN 343 374  
FT DOMAIN 374 405  
FT DOMAIN 405 436  
FT DOMAIN 436 467  
FT DOMAIN 467 498  
FT DOMAIN 498 529  
FT DOMAIN 529 560  
FT DOMAIN 560 591  
FT DOMAIN 591 680  
FT DOMAIN 681 771  
FT DOMAIN 772 862  
FT DOMAIN 863 954  
FT DOMAIN 955 1042  
FT DOMAIN 1043 1133  
FT DOMAIN 1134 1224  
FT DOMAIN 1225 1315  
FT DOMAIN 1316 1404  
FT DOMAIN 1405 1492  
FT DOMAIN 1493 1580  
FT DOMAIN 1589 1808  
FT DISULFID 64 64  
FT DISULFID 132 202  
FT DISULFID 196 207  
FT DISULFID 209 218  
FT DISULFID 223 233  
FT DISULFID 227 238  
TENASCIN.  
COILED COIL (POTENTIAL).  
EGF-LIKE 1 (INCOMPLETE).  
EGF-LIKE 2.  
EGF-LIKE 3.  
EGF-LIKE 4.  
EGF-LIKE 5.  
EGF-LIKE 6.  
EGF-LIKE 7.  
EGF-LIKE 8.  
EGF-LIKE 9.  
EGF-LIKE 10.  
EGF-LIKE 11.  
EGF-LIKE 12.  
EGF-LIKE 13.  
EGF-LIKE 14.  
FIBRONECTIN TYPE-III 1.  
FIBRONECTIN TYPE-III 2.  
FIBRONECTIN TYPE-III 3.  
FIBRONECTIN TYPE-III 4.  
FIBRONECTIN TYPE-III 5.  
FIBRONECTIN TYPE-III 6.  
FIBRONECTIN TYPE-III 7.  
FIBRONECTIN TYPE-III 8.  
FIBRONECTIN TYPE-III 9.  
FIBRONECTIN TYPE-III 10.  
FIBRONECTIN TYPE-III 11.  
FIBRINOGEN BETA/GAMMA.  
INTERCHAIN (POTENTIAL).  
BY SIMILARITY.  
BY SIMILARITY.  
BY SIMILARITY.  
BY SIMILARITY.

FT DISULFID 240  
FT DISULFID 254  
FT DISULFID 258  
FT DISULFID 271  
FT DISULFID 285  
FT DISULFID 289  
FT DISULFID 302  
FT DISULFID 316  
FT DISULFID 320  
FT DISULFID 333  
FT DISULFID 347  
FT DISULFID 351  
FT DISULFID 364  
FT DISULFID 378  
FT DISULFID 382  
FT DISULFID 395  
FT DISULFID 409  
FT DISULFID 413  
FT DISULFID 426  
FT DISULFID 440  
FT DISULFID 444  
FT DISULFID 457  
FT DISULFID 471  
FT DISULFID 475  
FT DISULFID 488  
FT DISULFID 502  
FT DISULFID 506  
FT DISULFID 517  
FT DISULFID 519  
FT DISULFID 581  
FT CARBOHYD 38  
FT CARBOHYD 168  
FT CARBOHYD 186  
FT CARBOHYD 228  
FT CARBOHYD 303  
FT CARBOHYD 643  
FT CARBOHYD 751  
FT CARBOHYD 759  
FT CARBOHYD 1050  
FT CARBOHYD 1090  
FT CARBOHYD 1101  
FT CARBOHYD 1112  
FT CARBOHYD 1153  
FT CARBOHYD 1183  
FT CARBOHYD 1416  
FT CARBOHYD 1736  
FT CARBOHYD 1769  
FT VARSPLIC 1043  
FT VARSPLIC 1043  
FT CONFLICT 182  
FT CONFLICT 563  
FT CONFLICT 598  
FT CONFLICT 838  
FT CONFLICT 886  
SQ SEQUENCE 1808 AA; 198858 MW; B924A06CF9EFD6DE CRC64;  
249 BY SIMILARITY.  
264 BY SIMILARITY.  
269 BY SIMILARITY.  
280 BY SIMILARITY.  
285 BY SIMILARITY.  
300 BY SIMILARITY.  
311 BY SIMILARITY.  
326 BY SIMILARITY.  
331 BY SIMILARITY.  
334 BY SIMILARITY.  
357 BY SIMILARITY.  
362 BY SIMILARITY.  
373 BY SIMILARITY.  
388 BY SIMILARITY.  
393 BY SIMILARITY.  
404 BY SIMILARITY.  
419 BY SIMILARITY.  
424 BY SIMILARITY.  
435 BY SIMILARITY.  
450 BY SIMILARITY.  
455 BY SIMILARITY.  
466 BY SIMILARITY.  
481 BY SIMILARITY.  
486 BY SIMILARITY.  
497 BY SIMILARITY.  
512 BY SIMILARITY.  
517 BY SIMILARITY.  
528 BY SIMILARITY.  
543 BY SIMILARITY.  
548 BY SIMILARITY.  
559 BY SIMILARITY.  
574 BY SIMILARITY.  
579 BY SIMILARITY.  
590 BY SIMILARITY.  
38 N-LINKED (GLCNAC. .) (POTENTIAL).  
168 N-LINKED (GLCNAC. .) (POTENTIAL).  
186 N-LINKED (GLCNAC. .) (POTENTIAL).  
228 N-LINKED (GLCNAC. .) (POTENTIAL).  
303 N-LINKED (GLCNAC. .) (POTENTIAL).  
643 N-LINKED (GLCNAC. .) (POTENTIAL).  
751 N-LINKED (GLCNAC. .) (POTENTIAL).  
759 N-LINKED (GLCNAC. .) (POTENTIAL).  
1050 N-LINKED (GLCNAC. .) (POTENTIAL).  
1090 N-LINKED (GLCNAC. .) (POTENTIAL).  
1101 N-LINKED (GLCNAC. .) (POTENTIAL).  
1112 N-LINKED (GLCNAC. .) (POTENTIAL).  
1153 N-LINKED (GLCNAC. .) (POTENTIAL).  
1183 N-LINKED (GLCNAC. .) (POTENTIAL).  
1416 N-LINKED (GLCNAC. .) (POTENTIAL).  
1736 N-LINKED (GLCNAC. .) (POTENTIAL).  
1769 N-LINKED (GLCNAC. .) (POTENTIAL).  
1043 MISSING (IN 200 KDA ISOFORM).  
1043 MISSING (IN 190 KDA ISOFORM).  
182 W -> R (IN REF. 2).  
563 SCPNDNNV -> PAPMTATTW (IN REF. 3).  
598 E -> G (IN REF. 3).  
838 T -> TEY (IN REF. 3).  
886 N -> F (IN REF. 3).  
SQ SEQUENCE 1808 AA; 198858 MW; B924A06CF9EFD6DE CRC64;

## Query Match

Best Local Similarity 5.0%; Score 732.5; DB 1; Length 1808;  
Matches 399; Conservative 241; Mismatches 716; Indels 543; Gaps 89;

QY 535 CPRNCHNGECVSGVCHCFPGFLGADCAKACPVLCSGNGQYSGKTCOCYSGWKGAECDDV 594  
DB 316 CPNDCFDRCRCINGTCFCEEGYTGDCGELTCPNKNCNGRCNGELCVCHGCFVGDGCS- 374  
QY 595 PMNQCIDPSCGGHSCIDGNCVCSAGYKGEHCEEVDCLDPTCCSHGVGVNCGELCSPGWG 654  
DB 375 -QKRC-PKDCNNRGRHCVDRGRCVCHGELGDCGELRCPN-DCHNRGRGTCINGQCVDEGFI 431





FT DISULFID 501 BY SIMILARITY. 511  
FT DISULFID 505 BY SIMILARITY. 516  
FT DISULFID 518 BY SIMILARITY. 527  
FT DISULFID 532 BY SIMILARITY. 542  
FT DISULFID 536 BY SIMILARITY. 547  
FT DISULFID 549 BY SIMILARITY. 558  
FT DISULFID 563 BY SIMILARITY. 573  
FT DISULFID 567 BY SIMILARITY. 578  
FT DISULFID 580 BY SIMILARITY. 589  
FT DISULFID 594 BY SIMILARITY. 604  
FT DISULFID 598 BY SIMILARITY. 609  
FT DISULFID 611 BY SIMILARITY. 620  
FT CARBOHYD 38 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 166 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 184 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 327 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 788 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1018 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1034 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1079 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1093 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1119 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1184 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1210 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1261 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1275 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1301 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1366 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1392 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1445 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1455 N-LINKED (GLCNAC. .) (POTENTIAL).  
Query Match 4.7%; Score 680; DB 1; Length 2201;  
Best Local Similarity 20.3%; Pred. No. 6.2e-29;  
Matches 454; Conservative 263; Mismatches 789; Indels 726; Gaps 108;  
QY 535 CPRNCHNGECVSGVCHFCFGFLGADCAKAAACPVLC- GNGOYKSGTCQCYSGWKGAED 593  
DB 221 CPSCDQGGKCVNGVGCFCFGAGADCSREICPVPCSEHGTCVGLCVCHDGFAGDDCN 280  
QY 594 VPMNQCIDPSCGGHSGIDGNCVCSAGYKGEHCBEVCDLPDTCSSHGVCVNGECLCSPGW 653  
DB 281 KPL--CLN-NCYNRGRVENECVDEGFTGEDCSELLCPN-DCFDGRRCINGTCYCEBGF 336  
QY 654 GGLNCELARVOCPCQCSGHGTYLPDTGLCSDCPNMWPGDCSVEVCVSDCETHGVICGAC 713  
337 TGEDC--GKPTCPHACTQGR--EEGQCVCDGFAGLDCSEKRCPADCHNRGRVCVDRG 392  
QY 714 RCEGHTGAACDQVCHPRCIEHCTCKDKCEGREGNHECHTIGR-----QTAGTETDG 768  
DB 393 ECDGFTGADCGELKCPNGCSGHGRVCVNGQCVCDGTYGDCSOLRCPNDCHSRGRVCG 452  
QY 769 -----CPDLNNGRGTCLGONSMQCVCTGWRGPGCNVAMETSCADNK 811  
DB 453 KVCCEQGFKGVCDSMPCNDCHQGRVCNG-----MCVDDGTYGEDCR---DRQCPDRC 505  
QY 812 DNEG--DGLVDCLD---PDC-----CL--QSAQNSLL---CRGSRDPL 845  
DB 506 SNRGLCVGQCVCDGFTGPDCAELSCPNCHGQGRVCVNGQCVCHGFGMKDCKEQRCPS 565  
QY 846 DIIQOQO-TWPAV--KSFYDRIKLAGKD-STHIIPEGNFNSLSVLIRGQVVTDTGT 901  
DB 566 DCHGQGRVDCQICHEGF-----TGLDGOHQSCPSD---CNNLQCVSGRGCICNEG- 614  
QY 902 PLVGVNVSFVKYKGYTIRQDGTFDLIANGGASLTLHFERAPFMQERTVWLPW-NSF 960  
DB 615 -YSGEDCSEVSPK-----DLVW---TEVT-----EETVNLAWDNEM 647  
QY 961 YAMDTLVKMEENSIPSCDLSGFRVDPPIIISPLSTFFSAAPGONPIVPTQVLHHEIE 1020  
DB 648 RVTEVLVYTPTH-----EGGLEMQF-RVPGD-----QTSTIIIELE 683  
QY 1021 LPSGNKRLYLSRTAGYKSLKTKMTQSTVPLNLRVHLMAVEGHLFOKSFQASPNLA 1080

DB 684 -PGVEYFIR-----VFALENNKSPVSARVATVLPAGEGLKFKSIKETSVEVE 731  
QY 1081 STFIWDKTD-----AYQORVYGLSDAVSVGGEYE 1110  
DB 732 ---WDPDLIAFETWEIIFRNMKKEDEGETKSURRPTSRYOT--GLAP-----GOEYE 780  
QY 1111 -----TCPS-----LILWEKRTALQGFELDPNSL 1135  
DB 781 ISLHIVKNNTRPGCLKRVTTITRLDAPSQIEVKQVDTTALITWFKPLAIDGIEL----- 835  
QY 1136 GWSLDKHHILNVKSG--ILHKGTEGNOFLTQOAPITTSIGN--GRRRSISPCSNGL 1190  
DB 836 -----TYGKDVPGDRTTIDLTEDENOY-----SIGNLKPDTEVEVSIISRGD 879  
QY 1191 AEGNKLAPVALAVGIDGSLVVGDFNIRIFPSPNNVTSILELRNKFKNHNPAAHYVL 1250  
DB 880 MSSNP--AKETFTGLDAP-----RNLRRVSQTDNSITL-----EWRNKAADISYRI 925  
QY 1251 AVDPVSGSLVSDTNSRRIRYRVKSLSGTKDLAGNSEYVAGTGEQCLPFDEARCGGGKAI 1310  
DB 926 KYAPISG-----GDHAEVDVPKSOQA----- 946  
QY 1311 DATLMSPRGIAVKNGLMYFVDATMIRKVDONGIISTLLSNDLTAVRPLSCDSSMDVAQ 1370  
DB 947 -TKTTLTGL---RPGTEYGIGVSAY-KEDKESNPATINAATELTDPKDLQVSETAETS- 1000  
QY 1371 VRLEWPTDLA-----VN-----PMDNSLYVLNNVILRTTENHOVSIIAGR 1411  
DB 1001 LTLWKYPLAKFDYRLNYSLPTQWGVQVQLPRNTTSYVLRG---LEPGQYINVLTLAEK 1057  
QY 1412 PMHQVPCIDYSLSKLAHSAESAISAISHTGVL-----YITETDE-KKIN 1458  
DB 1058 GRHKSPARVKAASQA--PELENLTVEVGDGLRLNWTAAADQAYEHFIIQVQANKVE 1115  
QY 1459 RLROVTTNGEICLLAGASDCDCKDNVNCYSGDDAYATDAIINSPLSLAVAPD-GTIY 1517  
DB 1116 AARNLTVPGS--LRVADIPGLKAATPYTSIYVIGQYRT--PVLSEASTGETNLEGEW 1172  
QY 1518 IADLGNIRIRAVSNKPVNAFYEAASPGQEELY---VFENADGIHQYTVSLVTVGEVLY 1574  
DB 1173 VAEVG-----WDALKLNTAPEGAYEYFIQVQEA-----TVEAAQ 1209  
QY 1575 NITY-----STDNDVTELIDNNGSLK-IRRDSSGMRPHLLMPDNQIITLVTVNGGLKV 1628  
DB 1210 NLTVPGGLRSTDLPLGLKAATHYITIRGVTQDFSTPLSV-----EVLTEEPVDMGNLTV 1264  
QY 1629 --VSTQNLLEGLMTYDGNLTGLLATKSDETGWTTFYDYDHEGRLTNTRP----- 1675  
DB 1265 TEVSWDALRLNWTTPDGTQDFTIQVQEA-----DQVEEAHNLTVPGSLRSMEIPGL 1316  
QY 1676 ---TGVVTSLHREMEKSTITDIENSRDDDVIT-----NLSS 1711  
DB 1317 RAGTPYTVTLHGEVR-----GHSRPLAVEVVTEDLPQLGLDVLAVSEVGHGDLRLNWT 1369  
QY 1712 VEASYT--VVQDQVRNSYQLCNN---GTLRMVYANGM-----ISFSEPHV 1753  
DB 1370 ADNAYEHFVQVQEVKNVEAAQNLTLPGSLRAVDIPGLEAATPYRVSIVYIRGYRTPVL 1429  
QY 1754 LAGTIT--PTIGRCNIS-----LPMENGLNSIEWRL--- 1782  
DB 1430 SAEASTAKEPEIGNLVNSDITPESFNLSMWATDGFETFTTEIIDSNRLLETVEYNISGA 1489  
QY 1783 -RKEQIKGK-----VTIFGRKLVRHGRNL-----LSIDYDRN----- 1813  
DB 1490 ERTAHISGLPPSTDFIVYLSGLAPSIRTKTISATATTAATLPLENLTIS-DINYPGTVS 1548  
QY 1814 -IRTEKIYDD-----HRKFTLRIYDQVGRPFLMLPSGLAAV 1850  
DB 1549 WMASENAFDSFLVTVDSGKLLDPQEFTLSTQKRLERGLITIG-----YEV 1597  
QY 1851 NVSYFFNRLAGLAGWAMSERTIDKQGRIVSRMFADCKWVSYS-----YLDKSWLLQOS 1906

Db 1598 MVSFGTQGHOTKPLRAEIVTEAPEVDNLLVSDATPDGFRLSWTADGEGVDFNLKIRDT 1657  
QY 1907 QRQYIFEYDSSDRLLAVTPSVAHSMSTHTSIGYIRNIYNPPESNASVIFDYSDDGRLT 1966  
Db 1658 KQO-----SEPLEITLLAPETRLDTGLREATEVEIEIYGLSKGRSQTVS-----AI 1705  
QY 1967 KTSFLGTGRQYKYKGLSKLSEIYVDSTAVTFGYDE-----TTGVLKMWNL 2013  
Db 1706 ATTAMGSPKEVIF-----SDITENSATVSWRAPTAQVESFRITVYVITGTPSMVTV 1757  
QY 2014 QSGGSCITRYRKIGPLVDKQI-----YRFE-----EGWVAREDYTH 2053  
Db 1758 D--GKTKQTRKLKIPGVEIYLSIIAMKGFEESEPVSGFTTALDGPGLVTANI----- 1810  
QY 2054 DNSFRIASIKPVISPTPLVDLYRDEISCKVEHFKF---GVIVYDINOIITTAVMTLS 2110  
Db 1811 TQSEALARQPAIA---TVDSYVVISYGEKYPEITRTVSGNTVEYALTDLEPATEYTL- 1865  
QY 2111 KHFDPHGRK-----EVQEMFRLSMYMTVQVDSMGRVTKREL 2149  
Db 1866 RIFAEBKPGOKSITIAKFTTDLSDPRDLTATEVQSE--TALLTWRP----- 1909  
2150 KLGPVANTTKYDYDG--DQOLQSAVANDRPTWRYSYDLNGLHLLNPGNSVRLMPLRYD 2208  
Db 1910 ---PRASVTGYLLVYESVDGKVEIVGPDTSYSIADLSPSTHY-----TAKIQALNGP 1961  
QY 2209 LRDR1-----TRLGDVQYKIDDDGYLCQSGDIFEYNSKGLLTRYN----- 2250  
Db 1962 LRSNMQITFTTIG-LLYPFPKDCSQAMLNGD---TTSGLYTIYVNGKAQALEVFCDM 2016  
QY 2251 ---KASGWSVQYRVDGVR-----ASYKTNLGHILQYPSYDLNPTRTTHVYNNHNSI 2302  
Db 2017 TSDGGGWIYFLRK-N-GRENFYQNKAYAAAGFGDRREFWLGDLNKLKITAQGY----- 2070  
QY 2303 TSLYLDLOGLHFAMESSSEEEYVVASDNTGTGPLAYSINGLMK-OLQYTAIGEYIDSN 2361  
Db 2071 -ELRVDLRH-----GETAFAYDK-----FSVGDAKTRYKLKRVGYSGTAGDS- 2113  
QY 2362 PDFQMVIGPHGG 2373  
Db 2114 -----MAYHNG 2119  
RESULT 3  
TENA\_PIG STANDARD; PRT: 1746 AA.  
AC Q29116; P98142;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
Tenaascin precursor (TN) (Hexabrachion) (Cytotactin) (Neuronectin)  
(GMBM) (Jf) (Miotendinous antigen) (Glioma-associated-extracellular  
matrix antigen) (GP 150-225) (Tenaascin-C) (TN-C) (P230).  
GN HXB.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Submaxillary gland;  
RX MEDLINE=92104189; PubMed=1722152;  
RA Nishi T., Weinstein J., Gillespie W.M., Paulson J.C.;  
RT "Complete primary structure of porcine tenascin: detection of  
tenascin transcript in adult submaxillary glands.";  
RL Eur. J. Biochem. 202:643-648(1991).  
RN [2]  
RP SEQUENCE OF 813-825; 887-917; 998-1011; 1597-1608 AND 1719-1730.  
RC TISSUE-Fetal brain;  
RX MEDLINE=98158323; PubMed=9498558;  
RA Wakatsuki S., Ho S.H., Arioka M., Yamasaki M., Kitamoto K.;  
RT "Isolation and characterization of a 230 kDa protein (p230)  
specifically expressed in fetal brains: its involvement in neurite

outgrowth from rat cerebral cortex neurons grown on monolayer of  
astrocytes.";  
J. Biochem. 122:1146-1152(1997).  
- FUNCTION: SAM (SUBSTRATE-ADHESION MOLECULE) THAT APPEARS TO  
INHIBIT CELL MIGRATION. MAY PLAY A ROLE IN SUPPORTING THE GROWTH  
OF EPITHELIAL TUMORS. IS A LIGAND FOR INTEGRINS ALPHA-8/BETA-1,  
ALPHA-9/BETA-1, ALPHA-V/BETA-3 AND ALPHA-V/BETA-6.  
- FUNCTION: PLAYS A ROLE DURING EARLY BRAIN DEVELOPMENT PARTICULARLY  
IN GROWTH CONE GUIDANCE. INVOLVED IN NEURITE OUTGROWTH FROM  
CORTICAL NEURONS GROWN ON THE MONOLAYER OF ASTROCYTES.  
- SUBUNIT: HEXAMERIC. A HOMOTRIMER MAY BE FORMED IN THE TRIPLE  
COILED-COIL REGION AND MAY BE STABILIZED BY DISULFIDE RINGS AT  
BOTH ENDS. TWO OF SUCH HALF-HEXABRACHIONS MAY BE DISULFIDE LINKED  
WITHIN THE CENTRAL GLOBULE.  
- SUBCELLULAR LOCATION: Extracellular matrix.  
- ALTERNATIVE PRODUCTS: 3 ISOFORMS: MAJOR, MINOR-1 AND MINOR-2  
(SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING. THEY ARE  
PRODUCED IN A TISSUE-AND TIME-SPECIFIC MANNER DURING DEVELOPMENT.  
- TISSUE SPECIFICITY: SUBMAXILLARY GLANDS AND BRAIN.  
- DEVELOPMENTAL STAGE: PREDOMINANTLY EXPRESSED IN THE EMBRYONIC AND  
EARLY POSTNATAL STAGES. LITTLE OR NO DETECTION IN ADULT BRAIN.  
- INDUCTION: BY TGF-BETA.  
- SIMILARITY: CONTAINS 15 EGF-LIKE DOMAINS.  
- SIMILARITY: CONTAINS 11 FIBRONECTIN TYPE III-LIKE DOMAINS.  
- SIMILARITY: CONTAINS 1 FIBRONECTIN C-TERMINAL DOMAIN.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: X61599; CAA43796.1; -  
HSSP: P24821; ITEN.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR003961; FN\_III.  
DR InterPro: IPR002181; Fibrinogen\_C.  
DR InterPro: IPR003962; FnIII\_repeat.  
DR InterPro: IPR002049; Laminin\_EGF.  
DR Pfam: PF00008; EGF\_14.  
DR Pfam: PF00147; fibrinogen\_C; 1.  
DR Pfam: PF00041; fn3; 10.  
DR PRINTS: PR00011; EGF\_LAMININ.  
DR PRINTS: PR00014; FNTYPEIII.  
DR SMART: SM00181; EGF; 9.  
DR SMART: SM00001; EGF\_Like; 2.  
DR SMART: SM00186; FBG; 1.  
DR SMART: SM00060; FN3; 9.  
DR PROSITE: PS00022; EGF\_1; 15.  
DR PROSITE: PS01186; EGF\_2; 14.  
KW Glycoprotein; Cell adhesion; Repeat; EGF-like domain; Coiled coil;  
KW Extracellular matrix; Alternative splicing; Signal.  
FT SIGNAL 1 22  
FT CHAIN 23 1746  
FT DOMAIN 23 185 INVOLVED IN HEXAMER FORMATION.  
FT DOMAIN 118 145 COILED COIL (POTENTIAL).  
FT DOMAIN 174 186 EGF-LIKE 1 (INCOMPLETE).  
FT DOMAIN 187 217 EGF-LIKE 2.  
FT DOMAIN 218 249 EGF-LIKE 3.  
FT DOMAIN 250 280 EGF-LIKE 4.  
FT DOMAIN 281 311 EGF-LIKE 5.  
FT DOMAIN 312 342 EGF-LIKE 6.  
FT DOMAIN 343 373 EGF-LIKE 7.  
FT DOMAIN 374 404 EGF-LIKE 8.  
FT DOMAIN 405 435 EGF-LIKE 9.  
FT DOMAIN 436 466 EGF-LIKE 10.  
FT DOMAIN 467 497 EGF-LIKE 11.  
FT DOMAIN 498 528 EGF-LIKE 12.  
FT DOMAIN 529 559 EGF-LIKE 13.  
FT DOMAIN 560 589 EGF-LIKE 14.  
FT DOMAIN 590 620 EGF-LIKE 15.

```
FT DOMAIN 621 710 FIBROECTIN TYPE-III 1.
FT DOMAIN 711 801 FIBROECTIN TYPE-III 2.
FT DOMAIN 802 891 FIBROECTIN TYPE-III 3.
FT DOMAIN 892 983 FIBROECTIN TYPE-III 4.
FT DOMAIN 984 1071 FIBROECTIN TYPE-III 5.
FT DOMAIN 1072 1162 FIBROECTIN TYPE-III 6.
FT DOMAIN 1163 1253 FIBROECTIN TYPE-III 7.
FT DOMAIN 1254 1343 FIBROECTIN TYPE-III 8.
FT DOMAIN 1343 1430 FIBROECTIN TYPE-III 9.
FT DOMAIN 1431 1518 FIBROECTIN TYPE-III 10.
FT DOMAIN 1527 1733 FIBRINOGEN C-TERMINAL.
FT DISULFID 64 64 INTERCHAIN (POTENTIAL).
FT DISULFID 190 200 BY SIMILARITY.
FT DISULFID 194 205 BY SIMILARITY.
FT DISULFID 207 216 BY SIMILARITY.
FT DISULFID 221 231 BY SIMILARITY.
FT DISULFID 225 236 BY SIMILARITY.
FT DISULFID 238 247 BY SIMILARITY.
FT DISULFID 252 263 BY SIMILARITY.
FT DISULFID 256 268 BY SIMILARITY.
FT DISULFID 270 279 BY SIMILARITY.
FT DISULFID 284 294 BY SIMILARITY.
FT DISULFID 288 299 BY SIMILARITY.
FT DISULFID 301 310 BY SIMILARITY.
FT DISULFID 315 325 BY SIMILARITY.
FT DISULFID 319 330 BY SIMILARITY.
FT DISULFID 332 341 BY SIMILARITY.
FT DISULFID 346 356 BY SIMILARITY.
FT DISULFID 350 361 BY SIMILARITY.
FT DISULFID 363 372 BY SIMILARITY.
FT DISULFID 377 387 BY SIMILARITY.
FT DISULFID 387 392 BY SIMILARITY.
FT DISULFID 394 403 BY SIMILARITY.
FT DISULFID 408 418 BY SIMILARITY.
FT DISULFID 412 423 BY SIMILARITY.
FT DISULFID 425 434 BY SIMILARITY.
FT DISULFID 439 449 BY SIMILARITY.
FT DISULFID 454 465 BY SIMILARITY.
FT DISULFID 470 480 BY SIMILARITY.
FT DISULFID 474 485 BY SIMILARITY.
FT DISULFID 487 496 BY SIMILARITY.
FT DISULFID 501 511 BY SIMILARITY.
FT DISULFID 505 516 BY SIMILARITY.
FT DISULFID 518 527 BY SIMILARITY.
FT DISULFID 532 542 BY SIMILARITY.
FT DISULFID 536 547 BY SIMILARITY.
FT DISULFID 549 558 BY SIMILARITY.
FT DISULFID 563 573 BY SIMILARITY.
FT DISULFID 567 578 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 594 604 BY SIMILARITY.
FT DISULFID 598 609 BY SIMILARITY.
FT DISULFID 611 620 BY SIMILARITY.
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 166 166 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 184 184 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 327 327 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 788 788 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1034 1034 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1079 1079 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1121 1121 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1354 1354 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 1072 1162 MISSING (IN MAJOR ISOFORM).
FT CONFLICT 1007 1007 T -> M (IN REF. 2).
SQ SEQUENCE 1746 AA: 191399 MW: 56349BICFE5E3C88 CRC64;
```

Query Match 4.5%; Score 656; DB 1; Length 1746;

Best Local Similarity 21.0%; Pred. No. 8.9e-28;

Matches 379; Conservative 211; Mismatches 642; Indels 574; Gaps 84;

```
QY 535 CPNCHNGNECVSGVCHCFPGFLGADCAKACPVLCSNGQYSGTKCQCYSGMKGACEDV 594
DB 315 CPKDCDFGRGRCINGTCYCDGEFEGEDCGLRACPHGCRGRCEBQCVCDEGFAGADCS- 373
QY 595 PMNOCIDPS-CGGHGSIDGNCVCSAGYKGEHCEVEVDCLDPTCSSHGVCVNGECLCSPGW 653
DB 374 -ERRC--PSCDHNRRCLDGRCECDGDFEGEDGELRC-PGGCSGHGRCVNGQCVCDBGR 429
QY 654 GGLNCELARVOCPCQCSGHGTYLPDTGLCSDPDNMNMGPDSCSVESVDCGTHGVCIGAC 713
DB 430 TGEDC--SOLRCPNDCHGRGRCV--QGRCECEHGFQGYDCSEMSPHDCGHGRCVNGMC 485
QY 714 RCEEGWGAACDQRVCHPRCIEHGTCKDGKCEGNGEHCITIGROTAGTGTGCPDLAC 773
DB 486 VDDGYGTGEDCRELCPCDCSQRGRCVDRGRCVCEHGFAGPDC-----DLACPSDC 536
QY 774 NNGRCITLGNSSWOCVOTGWRGPGCNVAMETSCADNKNEDGLVCLCLDPC----- 826
DB 537 HGRGRCVNG----QCVCHGFTGKDCG--QRRCPG--DCHGQG--RCVDGQCVCHEGFT 585
QY 827 ---CLQSACQNSLLCRGSRDPLDIIQOQOTDPAVPAKSFYDRIKLLAGKDSHIIPGPNF 883
DB 586 GLDCGQRSCPN-----DCSNWQ----- 603
QY 884 NSSLVSLIRGOVTTDGTPLGVNVVFKYKPYGTYTITRODGTDLIANGASLTLPHER 943
DB 604 -----CVSGRCICNEG--YSGDCSQVSPK-----DLIV---TEVT----- 635
QY 944 APMSQERTVHLPW-NSFYAMDITLVKMTSENS-----IPSCDLSGFVRPDPITLISP 994
DB 636 -----EETVNLAWDNEMRVTEYLIVYTPTHEDGLEWQFRVPGDQSTTIR----- 680
QY 995 LSTFFSAAPGONPIVPTQVLEHEIELPGSNVKLRLSSRTAGYKSLKITMTQSTVPLN 1054
DB 681 -----ELEPGVEYIRVFAILENKKSP-----VSARVATY----- 711
QY 1055 LIRVHLMVAVEGHLPQKSFQASPNLSTFIWDKTDAYQORVYGLSDAVVSGFEYETPCS 1114
DB 712 -----LPTPEGLKFKSIKETSVEVE---WDPLD-----IAFE----- 740
QY 1115 LILWE-----KRTALQGFELDPNSLGNWGLDKHILNVK 1149
DB 741 --TWEIIFRNKNKEDEGEITKSLRRPETTYRQTGLAPQGEYEIS-----LHI--VK 787
QY 1150 SGILHKGTTGENQFLTQOPAIITSIMGNRRRSISCPs---CNGLAEGNKLILA---PVALA 1203
DB 788 NNT--RGPGLKRVITR-----LDAPSQIEAKQVDTTALTITWFKPLAEI 830
QY 1204 VGIDGSLYVGFNFIIRIFPSRNVTSILELRNKEFKHNNPAHKYLLAVDPVSGSLYVSD 1263
DB 831 DGIELTYGIKD-----VPGDRTT--IDLTHEENQVYS-----IGNL-KPD 866
QY 1264 TNSRRIRYVKSLGTDKLAGNSVAVGTGEOCLPFDEARCGGKAIDATLMSRPIAVD 1323
DB 867 TE-----TEVSLSIRRADMSN-----PAKE-----TFTTGLDAPRN----- 898
QY 1324 KNLMIYFVATMIRKVDQNGIISTLLGSNDLTAVRPLSCDSSMDVAQVRLWEPTDLAVNP 1383
DB 899 -----LRRISQT-----DMS-----ITLEM----- 913
QY 1384 MDNSLYVLNNVILRITENHOVSTIAGRPMHCQV-----PGIDYSL 1425
DB 914 -----RNGKAAADTYRIKYAPISGGD-HAEVEVPRSPOTTTKATLGLRPGTEYIG 964
QY 1426 KLAIHSALESASATAISHTGVLYITEDEKKINRLQVTTNGEICLLAGASDCCKNDV 1485
DB 965 VSAVKGDKESDPAT-----INAAATDLDPPKDFRVSLEKESLTLMLWTPLAKFD-RYRL 1017
QY 1486 NCNCSYSGDDA-----YATDAILNSPSSLAVAPDGTIYI-ADLGNIRIRAVSKNPKVLNA 1538
DB 1018 NYGLPSGPQVPEVQLPRNATSYILR---GLEPGQEYITILLTAEKGR-----HKSKPA--- 1065
QY 1539 FNQYEAASPGEQELYVFNADGIHQYTVSLVTGEYLYNFTYS-----TDNDVTELIDNN 1591
```

```

Db 1066 --RVKASTAGEPE-----IGNLSVSDITPE---SFLSMTWTEGAFETFTETIDSN 1112
Qy 1592 GNSLKRRSDGNPR--HL--LMPDQNIITLTVGTNGGL--KVVS-----TONLE 1635
Db 1113 RELETHEYNISGAERTAHISGLRPGNDFIYISGLAPGQTQKPSATATATEAPEVDNLL 1172
Qy 1636 LGLMTYDNGTGLATKSDETGWTTFY----DVDHEGRLTNVT-----RPTGVVTSLSHREME 1687
Db 1173 VSDATPDGFR--LSWTADEGVDFSEVLKIRDTKKQSEPLEITLLASERTDRITGLREATE 1230
Qy 1688 KSTTI--DIENSNRDDDVTVITNL---SSVEASYTVVQDVQRNSYQLCNNGTLRVMTANGM 1743
Db 1231 YEIELYGISSGRKSQVSAIATAMGSKPEITFSDITE-----NSATVSMWVPTAQ 1281
Qy 1744 GTSFH-----SEPHVLGATITPTIGRCNISLPMENGLNSTEWRLEQIKGVKTIFG 1795
Db 1282 VESFRITYPIVITGGARSVTVVDTGKTQTRLLRLLP-----GVEYLVSVIAVKG-----FE 1331
Qy 1796 RKLVRHGRMLLSIDYDRNTRTEKIYDDHRKFTLRITYDQVGRFLLWPSGLAAVNVSYF 1855
Db 1332 ESEPVSGTLTTALDGPGLVITANITDSEAL-----AMQPAIAPVDHYVISY 1378
1856 FNGRLAGLORGAMSER-----TDIDKQGRIVSRMFADGKVMWSYSLDKSMVLLLSQSORQY 1910
Db 1379 TGDVRPEITRTVSGNTVEVALINLEPATEYTLRIFAEKGPKQKSTTTFTTDDLSPRDL 1438
Qy 1911 -IFEYDSSDRLLAVTMP--SVARHSMSTHTSIGIYRIINYPPESSNASVIFDYSDDGRIKL 1967
Db 1439 TATEVQSEATALLTWPRASVTVGLLVYESVDGTLKEVVVGPETTS-----YSLSGLSFS 1493
Qy 1968 TSFLGTGRGVFKYKGLSKLSEIVYDSTAVTFGY-----DETGVLMK-----2010
Db 1494 THY--TARIQALNGPLRSKMSQTVFTTIGLXPFPFRDCSQAMLNGTTSGLTYIYVNDK 1551
Qy 2011 -----VNLQSGGFSECTIR-----YRK-----IGPLVDK-----QIYRFSE 2040
Db 1552 AOKLEVFCDMTSDSGGWIIVFLRRKNGREDFYRNWKAAYAAGFDLKEEFWGLGDALSKIYA 1611
Qy 2041 EGMVNAFDTYTHDNS-----FRIASIKPVISETPLPVDLYRYDEISGKVEHFGK-FGVITY 2095
Db 1612 QGOYELRVLDRHGEYAYAVYDRFSVGDARTYKLVKEGYSCTAGDSMAYHNGRSESTFD 1671
Qy 2096 YDINQIITAVMTLSKHF-----DTHGRKEVOYEMFRSLMYMTVQYDSMGRIKVR 2147
Db 1672 KDTDSAITNCALSYKGAFWYKNCRVNLMGRYGDNSHSGQVWFMFWKKGHEYS-----IQFA 1727
Qy 2148 ELKLG 2153
Db 1728 ENKLRP 1733
RESULT 4
TENX HUMAN
ID TENX HUMAN STANDARD; PRT: 4289 AA.
AC P22105; P78530; P78531; Q08424; Q0UMG7;
DT 01-AUG-1991 (Rel. 19, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Tenascin-X precursor (TN-X) (Hexabrachion-like).
GN TNXB OR TNX OR XB OR HXBL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowen L., Dankers C., Baskin D., Faust J., Loretz C., Ahearn M.E.,
RA Banca A., Schwartzell S., Smith T.M., Spies T., Hood L.;
RT "Sequence determination of 300 kilobases of the human class III MHC
RT locus."
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]

```

```

RP SEQUENCE OF 1-747 AND 1687-1944 FROM N.A.
RC TISSUE=Leukocyte;
RX MEDLINE=93300909; PubMed=7686164;
RA Bristow J., Tee M.K., Gitelman S.E., Mellon S.H., Miller W.L.;
RT "Tenascin-X: a novel extracellular matrix protein encoded by the human
RT XB gene overlapping P450c21b."
RL J. Cell Biol. 122:265-278(1993).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM XB-SHORT).
RC TISSUE=Adrenal gland;
RX MEDLINE=96015044; PubMed=8530023;
RA Tee M.K., Thomson A.A., Bristow J., Miller W.L.;
RT "Sequences promoting the transcription of the human XA gene
RT overlapping P450c21A correctly predict the presence of a novel,
RT adrenal-specific, truncated form of tenascin-X."
RL Genomics 28:171-178(1995).
RN [4]
RP SEQUENCE OF 1-23 FROM N.A.
RC TISSUE=Fetal adrenal gland;
RX MEDLINE=97081760; PubMed=8923003;
RA Speck M., Barry F., Miller W.L.;
RT "Alternate promoters and alternate splicing of human tenascin-X, a
RT gene with 5' and 3' ends buried in other genes."
RL Hum. Mol. Genet. 5:1749-1758(1996).
RN [5]
RP SEQUENCE OF 3470-4289 FROM N.A.
RX MEDLINE=89367293; PubMed=2475872;
RA Morel Y., Bristow J., Gitelman S.E., Miller W.L.;
RT "Transcript encoded on the opposite strand of the human steroid 21-
RT hydroxylase/complement component C4 gene locus."
RL Proc. Natl. Acad. Sci. U.S.A. 86:6582-6586(1989).
CC -!- FUNCTION: APPEARS TO MEDIATE INTERACTIONS BETWEEN CELLS AND THE
CC EXTRACELLULAR MATRIX. SUBSTRATE-ADHESION MOLECULE THAT APPEARS TO
CC INHIBIT CELL MIGRATION. MAY PLAY A ROLE IN SUPPORTING THE GROWTH
CC OF EPITHELIAL TUMORS.
CC -!- SUBCELLULAR LOCATION: Extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; XB (SHOWN HERE) AND
CC XB-SHORT; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN FETAL ADRENAL, IN FETAL
CC TESTIS, FETAL SMOOTH, STRIATED AND CARDIAC MUSCLE. XB-SHORT IS
CC ONLY EXPRESSED IN THE ADRENAL GLAND.
CC -!- DISEASE: ASSOCIATION WITH CONGENITAL ADRENAL HYPERPLASIA.
CC -!- DISEASE: Defects in TNXB are a cause of autosomal recessive
CC Ehlers-Danlos syndrome.
CC -!- SIMILARITY: CONTAINS 19 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 32 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS A DOMAIN RELATED TO THE C-TERMINAL PART OF
CC THE BETA AND GAMMA CHAINS OF FIBRINOGEN.
CC -!- CAUTION: THERE ARE TWO GENES FOR TN-X: TNXA AND TNXB. TNXA IS
CC A PARTIAL GENE WHICH CAN SOMETIME RECOMBINE WITH TNXB.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; U89337; AAB47488.1; -
DR EMBL; AF019413; AAB67981.1; -
DR EMBL; X71923; CAA50739.1; -
DR EMBL; Y13782; CAA74109.1; -
DR EMBL; Y13783; CAA74110.1; -
DR EMBL; U24488; AAB41287.1; -
DR EMBL; U52696; AAC50889.1; -
DR EMBL; M25813; AAA35884.1; -
DR HSSP; P02871; IFZD.
DR MIM; 600985; -
DR MIM; 606408; -
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR002181; Fibrinogen_C.

```



DR InterPro; IPR002049; Laminin\_EGF.  
 DR Pfam; PF00147; fibronogen\_C; 1.  
 DR PRINTS; PR00041; fn3; 6.  
 DR PRINTS; PR00011; EGF\_LAMININ.  
 DR SMART; SM00181; EGF; 8.  
 DR SMART; SM00001; EGF\_Like; 1.  
 DR SMART; SM00186; FBG; 1.  
 DR SMART; SM00060; FN3; 31.  
 DR PROSITE; PS00022; EGF\_1; 18.  
 DR PROSITE; PS01186; EGF\_2; 19.  
 DR PROSITE; PS00514; FIBRINAG\_C\_DOMAIN; 1.  
 KW Glycoprotein; Cell adhesion; Repeat; EGF-like domain; Coiled coil;  
 KW Extracellular matrix; Alternative splicing; Signal.  
 FT SIGNAL 1 23  
 FT CHAIN 24 4289  
 FT DOMAIN 156 167  
 FT DOMAIN 202 213  
 FT DOMAIN 233 244  
 FT DOMAIN 264 275  
 FT DOMAIN 295 306  
 FT DOMAIN 326 337  
 FT DOMAIN 357 368  
 FT DOMAIN 388 399  
 FT DOMAIN 419 430  
 FT DOMAIN 450 461  
 FT DOMAIN 481 492  
 FT DOMAIN 512 523  
 FT DOMAIN 543 554  
 FT DOMAIN 574 585  
 FT DOMAIN 605 616  
 FT DOMAIN 636 647  
 FT DOMAIN 667 678  
 FT DOMAIN 703 714  
 FT DOMAIN 734 745  
 FT DOMAIN 792 803  
 FT DOMAIN 822 833  
 FT DOMAIN 901 912  
 FT DOMAIN 941 952  
 FT DOMAIN 1047 1058  
 FT DOMAIN 1149 1160  
 FT DOMAIN 1246 1257  
 FT DOMAIN 1348 1359  
 FT DOMAIN 1459 1470  
 FT DOMAIN 1561 1572  
 FT DOMAIN 1659 1670  
 FT DOMAIN 1756 1767  
 FT DOMAIN 1856 1867  
 FT DOMAIN 1962 1973  
 FT DOMAIN 2069 2080  
 FT DOMAIN 2167 2178  
 FT DOMAIN 2266 2277  
 FT DOMAIN 2365 2376  
 FT DOMAIN 2471 2482  
 FT DOMAIN 2582 2593  
 FT DOMAIN 2688 2699  
 FT DOMAIN 2794 2805  
 FT DOMAIN 2889 2900  
 FT DOMAIN 2997 3008  
 FT DOMAIN 3105 3116  
 FT DOMAIN 3211 3222  
 FT DOMAIN 3307 3318  
 FT DOMAIN 3399 3410  
 FT DOMAIN 3494 3505  
 FT DOMAIN 3601 3612  
 FT DOMAIN 3699 3710  
 FT DOMAIN 3801 3812  
 FT DOMAIN 3890 3901  
 FT DOMAIN 3971 3982  
 FT DOMAIN 3978 3989  
 FT DOMAIN 4059 4070  
 FT DOMAIN 4071 4082  
 FT SITE 1748  
 FT DISULFID 4075  
 FT DISULFID 4227  
 FT CARBOHYD 31 32  
 FT CARBOHYD 901 902

FT CARBOHYD 930 930 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 3900 3900 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 3953 3953 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 3965 3965 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 4140 4140 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 1 3616 MISSING (IN ISOFORM XB-SHORT).  
 FT CONFLICT 135 135 G -> G (IN REF. 3 AND 5).  
 FT CONFLICT 4038 4038 P -> G (IN REF. 3 AND 5).  
 FT CONFLICT 4163 4163 M -> I (IN REF. 3 AND 5).  
 SQ SEQUENCE 4289 AA; 464454 MW; 84657C12C65C6470 CRC64;  
 Query Match 4.2%; Score 616; DB 1; Length 4289;  
 Best Local Similarity 30.5%; Pred. No. 5.8e-25;  
 Matches 153; Conservative 56; Mismatches 186; Indels 106; Gaps 19;  
 QY 532 VQDCPRNCHNGECVSGVCHCFPGFLGADCAKACAPVLCSSNGQYSGKTCQCYSGWKGA 591  
 DB 401 VRSPGDCNQGRGECDEGRVCVPGYTGDCGSRACPRDCRGRCENGVCVCNAGYSGED 460  
 QY 592 CDVPMNOCIDPSCGGHSCIDGNCVCSAGYKGEHCEVDCLDPTCSSHGVNCGECLSP 651  
 DB 461 CGV---RSCPDGCRGRGRCESGRCMWPYTGRCGTAC--PGDGRGRGRCVDRGCVNP 516  
 QY 652 GWGLNCELARVQCPDQCSHGHTYLPDTGLCSDPNMMPDCSVESVDCGTHGVCI 711  
 DB 517 GFTGEDCGSRR--CPGDCRHHG--LCEDGVCDAGYSGEDCSTRSCPGCGRGCQCLDG 572  
 QY 712 ACREEGHTGAACDQVRVCHPRCIEHGTCCKGCKECRENGEHCHTIGRTAGTETDGP 771  
 DB 573 RVCEDGYSGEDCGVROCPCNDCSQHGVCQGVGCICWEGYVSEDCSIRT-----CPS 623  
 QY 772 LCNGNGRCTLQNSWQVCQGTGWRGPGCNVAMETSC-ADNKD-----NEGDLV 819  
 DB 624 NCHGRGCEG-----RCLCDPGYTGPTCATRM---CPADCRGRGRCVQVCLCHGVYGE 676  
 QY 820 DC-----LDPCCLQSAQNSLLRG-----840  
 DB 677 DCGOEEPPASACPGCGCPRELCRAGQCVCVEGFRGPDCAIQT-CPGD--CRGRGECHDGS-733  
 QY 841 --SRPLDIIQGGTDPAPVKSFYDRIKLLAGKQ---STHIIIGEN---PENSILVSLIR 892  
 DB 734 CVCKDGVAGEDCGEARVPSSASAYDQRLAPQEQYQVTVRALRGTSWGLPASKITITMID 793  
 QY 893 G----QVVTDTGT-----PLGVNVSVFYKPKYGTITR-----QDGTFF--DLIAN 932  
 DB 794 GPQDLRVVAVTPTTLELGNLRPQAEVDREVVSVYSGNQVRVLEVPPEADCTLLTDLMPG 853  
 QY 933 GGASLTLHFERAPFMSQRTV 953  
 DB 854 VEYVTVTAERGRAVSPASV 874  
 RESULT 5  
 ID NTC1\_MOUSE STANDARD; PRT; 2531 AA.  
 AC Q01705;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Neurogenic locus notch homolog protein 1 precursor (NOTCH protein).  
 GN NOTCH1 OR MOTCH.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=93194170; PubMed=8449489;  
 RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,  
 RA Copeland N.G., Gridley T.;  
 RA "Cloning, analysis, and chromosomal localization of Notch-1, a mouse

RT homolog of Drosophila Notch.";  
 RL Genomics 15:259-264(1993).  
 [2]  
 RN SEQUENCE OF 1551-2170 FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=93048835; PubMed=1425352;  
 RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,  
 Greenspan R.J., McMahon A.P., Gridley T.;  
 RA "Expression pattern of Notch, a mouse homolog of Drosophila Notch,  
 RT suggests an important role in early postimplantation mouse  
 RL development.";  
 RL Development 115:737-744(1992).  
 CC -!- SUBCELLULAR LOCATION: type I membrane protein.  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.  
 CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.  
 CC -!- SIMILARITY: CONTAINS 5 ANK REPEATS.  
 CC -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z11886; CAA77941.1; -;  
 DR HSP; P00740; 1EDM.  
 DR MGD; MGI:97363; Notch1.  
 DR InterPro; IPR002110; ANK.  
 DR InterPro; IPR000152; Asx\_hydroxyl.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR000742; EGF 2.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR001438; EGF\_II.  
 DR InterPro; IPR000800; Notch.  
 DR Pfam; PF00023; ank; 6.  
 DR Pfam; PF00008; EGF; 35.  
 DR Pfam; PF00066; notch; 3.  
 DR PRINTS; PR00010; EGFBLD.  
 DR PRINTS; PR01452; NOTCH.  
 DR SMART; SM00248; ANK; 3.  
 DR SMART; SM00179; EGF\_CA; 23.  
 DR SMART; SM00001; EGF\_Like; 11.  
 DR SMART; SM00004; NL; 2.  
 DR PROSITE; PS50088; ANK\_REPEAT; 2.  
 DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 22.  
 DR PROSITE; PS00022; EGF\_1; 34.  
 DR PROSITE; PS01186; EGF\_2; 27.  
 DR PROSITE; PS01187; EGF\_CA; 21.  
 KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;  
 KW Transmembrane; Signal; Glycoprotein.  
 FT SIGNAL 1 18  
 FT CHAIN 19 2531  
 FT DOMAIN 19 1725  
 FT TRANSMEM 1726 1746  
 FT DOMAIN 1747 2531  
 FT DOMAIN 20 58  
 FT DOMAIN 59 99  
 FT DOMAIN 102 139  
 FT DOMAIN 140 176  
 FT DOMAIN 178 216  
 FT DOMAIN 218 255  
 FT DOMAIN 257 293  
 FT DOMAIN 295 333  
 FT DOMAIN 335 371  
 FT DOMAIN 372 410  
 FT DOMAIN 412 450  
 FT DOMAIN 452 488  
 FT DOMAIN 490 526  
 FT DOMAIN 528 564

566  
 603  
 641  
 676  
 714  
 716  
 751  
 789  
 827  
 867  
 905  
 943  
 981  
 1019  
 1057  
 1095  
 1143  
 1181  
 1219  
 1265  
 1305  
 1346  
 1384  
 1426  
 1462  
 1480  
 1522  
 1562  
 1597  
 1649  
 1683  
 1721  
 1759  
 1797  
 1833  
 1869  
 1905  
 1941  
 1977  
 2013  
 2049  
 2085  
 2121  
 2157  
 2193  
 2229  
 2265  
 2301  
 2337  
 2373  
 2409  
 2445  
 2481  
 2517  
 2553  
 2589  
 2625  
 2661  
 2697  
 2733  
 2769  
 2805  
 2841  
 2877  
 2913  
 2949  
 2985  
 3021  
 3057  
 3093  
 3129  
 3165  
 3201  
 3237  
 3273  
 3309  
 3345  
 3381  
 3417  
 3453  
 3489  
 3525  
 3561  
 3597  
 3633  
 3669  
 3705  
 3741  
 3777  
 3813  
 3849  
 3885  
 3921  
 3957  
 3993  
 4029  
 4065  
 4101  
 4137  
 4173  
 4209  
 4245  
 4281  
 4317  
 4353  
 4389  
 4425  
 4461  
 4497  
 4533  
 4569  
 4605  
 4641  
 4677  
 4713  
 4749  
 4785  
 4821  
 4857  
 4893  
 4929  
 4965  
 5001  
 5037  
 5073  
 5109  
 5145  
 5181  
 5217  
 5253  
 5289  
 5325  
 5361  
 5397  
 5433  
 5469  
 5505  
 5541  
 5577  
 5613  
 5649  
 5685  
 5721  
 5757  
 5793  
 5829  
 5865  
 5901  
 5937  
 5973  
 6009  
 6045  
 6081  
 6117  
 6153  
 6189  
 6225  
 6261  
 6297  
 6333  
 6369  
 6405  
 6441  
 6477  
 6513  
 6549  
 6585  
 6621  
 6657  
 6693  
 6729  
 6765  
 6801  
 6837  
 6873  
 6909  
 6945  
 6981  
 7017  
 7053  
 7089  
 7125  
 7161  
 7197  
 7233  
 7269  
 7305  
 7341  
 7377  
 7413  
 7449  
 7485  
 7521  
 7557  
 7593  
 7629  
 7665  
 7701  
 7737  
 7773  
 7809  
 7845  
 7881  
 7917  
 7953  
 7989  
 8025  
 8061  
 8097  
 8133  
 8169  
 8205  
 8241  
 8277  
 8313  
 8349  
 8385  
 8421  
 8457  
 8493  
 8529  
 8565  
 8601  
 8637  
 8673  
 8709  
 8745  
 8781  
 8817  
 8853  
 8889  
 8925  
 8961  
 8997  
 9033  
 9069  
 9105  
 9141  
 9177  
 9213  
 9249  
 9285  
 9321  
 9357  
 9393  
 9429  
 9465  
 9501  
 9537  
 9573  
 9609  
 9645  
 9681  
 9717  
 9753  
 9789  
 9825  
 9861  
 9897  
 9933  
 9969  
 10005  
 10041  
 10077  
 10113  
 10149  
 10185  
 10221  
 10257  
 10293  
 10329  
 10365  
 10401  
 10437  
 10473  
 10509  
 10545  
 10581  
 10617  
 10653  
 10689  
 10725  
 10761  
 10797  
 10833  
 10869  
 10905  
 10941  
 10977  
 11013  
 11049  
 11085  
 11121  
 11157  
 11193  
 11229  
 11265  
 11301  
 11337  
 11373  
 11409  
 11445  
 11481  
 11517  
 11553  
 11589  
 11625  
 11661  
 11697  
 11733  
 11769  
 11805  
 11841  
 11877  
 11913  
 11949  
 11985  
 12021  
 12057  
 12093  
 12129  
 12165  
 12201  
 12237  
 12273  
 12309  
 12345  
 12381  
 12417  
 12453  
 12489  
 12525  
 12561  
 12597  
 12633  
 12669  
 12705  
 12741  
 12777  
 12813  
 12849  
 12885  
 12921  
 12957  
 12993  
 13029  
 13065  
 13101  
 13137  
 13173  
 13209  
 13245  
 13281  
 13317  
 13353  
 13389  
 13425  
 13461  
 13497  
 13533  
 13569  
 13605  
 13641  
 13677  
 13713  
 13749  
 13785  
 13821  
 13857  
 13893  
 13929  
 13965  
 14001  
 14037  
 14073  
 14109  
 14145  
 14181  
 14217  
 14253  
 14289  
 14325  
 14361  
 14397  
 14433  
 14469  
 14505  
 14541  
 14577  
 14613  
 14649  
 14685  
 14721  
 14757  
 14793  
 14829  
 14865  
 14901  
 14937  
 14973  
 15009  
 15045  
 15081  
 15117  
 15153  
 15189  
 15225  
 15261  
 15297  
 15333  
 15369  
 15405  
 15441  
 15477  
 15513  
 15549  
 15585  
 15621  
 15657  
 15693  
 15729  
 15765  
 15801  
 15837  
 15873  
 15909  
 15945  
 15981  
 16017  
 16053  
 16089  
 16125  
 16161  
 16197  
 16233  
 16269  
 16305  
 16341  
 16377  
 16413  
 16449  
 16485  
 16521  
 16557  
 16593  
 16629  
 16665  
 16701  
 16737  
 16773  
 16809  
 16845  
 16881  
 16917  
 16953  
 16989  
 17025  
 17061  
 17097  
 17133  
 17169  
 17205  
 17241  
 17277  
 17313  
 17349  
 17385  
 17421  
 17457  
 17493  
 17529  
 17565  
 17601  
 17637  
 17673  
 17709  
 17745  
 17781  
 17817  
 17853  
 17889  
 17925  
 17961  
 18000  
 18036  
 18072  
 18108  
 18144  
 18180  
 18216  
 18252  
 18288  
 18324  
 18360  
 18396  
 18432  
 18468  
 18504  
 18540  
 18576  
 18612  
 18648  
 18684  
 18720  
 18756  
 18792  
 18828  
 18864  
 18900  
 18936  
 18972  
 19008  
 19044  
 19080  
 19116  
 19152  
 19188  
 19224  
 19260  
 19296  
 19332  
 19368  
 19404  
 19440  
 19476  
 19512  
 19548  
 19584  
 19620  
 19656  
 19692  
 19728  
 19764  
 19800  
 19836  
 19872  
 19908  
 19944  
 19980  
 20016  
 20052  
 20088  
 20124  
 20160  
 20196  
 20232  
 20268  
 20304  
 20340  
 20376  
 20412  
 20448  
 20484  
 20520  
 20556  
 20592  
 20628  
 20664  
 20700  
 20736  
 20772  
 20808  
 20844  
 20880  
 20916  
 20952  
 20988  
 21024  
 21060  
 21096  
 21132  
 21168  
 21204  
 21240  
 21276  
 21312  
 21348  
 21384  
 21420  
 21456  
 21492  
 21528  
 21564  
 21600  
 21636  
 21672  
 21708  
 21744  
 21780  
 21816  
 21852  
 21888  
 21924  
 21960  
 22000  
 22036  
 22072  
 22108  
 22144  
 22180  
 22216  
 22252  
 22288  
 22324  
 22360  
 22396  
 22432  
 22468  
 22504  
 22540  
 22576  
 22612  
 22648  
 22684  
 22720  
 22756  
 22792  
 22828  
 22864  
 22900  
 22936  
 22972  
 23008  
 23044  
 23080  
 23116  
 23152  
 23188  
 23224  
 23260  
 23296  
 23332  
 23368  
 23404  
 23440  
 23476  
 23512  
 23548  
 23584  
 23620  
 23656  
 23692  
 23728  
 23764  
 23800  
 23836  
 23872  
 23908  
 23944  
 23980  
 24016  
 24052  
 24088  
 24124  
 24160  
 24196  
 24232  
 24268  
 24304  
 24340  
 24376  
 24412  
 24448  
 24484  
 24520  
 24556  
 24592  
 24628  
 24664  
 24700  
 24736  
 24772  
 24808  
 24844  
 24880  
 24916  
 24952  
 24988  
 25024  
 25060  
 25096  
 25132  
 25168  
 25204  
 25240  
 25276  
 25312  
 25348  
 25384  
 25420  
 25456  
 25492  
 25528  
 25564  
 25600  
 25636  
 25672  
 25708  
 25744  
 25780  
 25816  
 25852  
 25888  
 25924  
 25960  
 25996  
 26032  
 26068  
 26104  
 26140  
 26176  
 26212  
 26248  
 26284  
 26320  
 26356  
 26392  
 26428  
 26464  
 26500  
 26536  
 26572  
 26608  
 26644  
 26680  
 26716  
 26752  
 26788  
 26824  
 26860  
 26896  
 26932  
 26968  
 27004  
 27040  
 27076  
 27112  
 27148  
 27184  
 27220  
 27256  
 27292  
 27328  
 27364  
 27400  
 27436  
 27472  
 27508  
 27544  
 27580  
 27616  
 27652  
 27688  
 27724  
 27760  
 27796  
 27832  
 27868  
 27904  
 27940  
 27976  
 28012  
 28048  
 28084  
 28120  
 28156  
 28192  
 28228  
 28264  
 28300  
 28336  
 28372  
 28408  
 28444  
 28480  
 28516  
 28552  
 28588  
 28624  
 28660  
 28696  
 28732  
 28768  
 28804  
 28840  
 28876  
 28912  
 28948  
 28984  
 29020  
 29056  
 29092  
 29128  
 29164  
 29200  
 29236  
 29272  
 29308  
 29344  
 29380  
 29416  
 29452  
 29488  
 29524  
 29560  
 29596  
 29632  
 29668  
 29704  
 29740  
 29776  
 29812  
 29848  
 29884  
 29920  
 29956  
 29992  
 30028  
 30064  
 30100  
 30136  
 30172  
 30208  
 30244  
 30280  
 30316  
 30352  
 30388  
 30424  
 30460  
 30496  
 30532  
 30568  
 30604  
 30640  
 30676  
 30712  
 30748  
 30784  
 30820  
 30856  
 30892  
 30928  
 30964  
 31000  
 31036  
 31072  
 31108  
 31144  
 31180  
 31216  
 31252  
 31288  
 31324  
 31360  
 31396  
 31432  
 31468  
 31504  
 31540  
 31576  
 31612  
 31648  
 31684  
 31720  
 31756  
 31792  
 31828  
 31864  
 31900  
 31936  
 31972  
 32008  
 32044  
 32080  
 32116  
 32152  
 32188  
 32224  
 32260  
 32296  
 32332  
 32368  
 32404  
 32440  
 32476  
 32512  
 32548  
 32584  
 32620  
 32656  
 32692  
 32728  
 32764  
 32800  
 32836  
 32872  
 32908  
 32944  
 32980  
 33016  
 33052  
 33088  
 33124  
 33160  
 33196  
 33232  
 33268  
 33304  
 33340  
 33376  
 33412  
 33448  
 33484  
 33520  
 33556  
 33592  
 33628  
 33664  
 33700  
 33736  
 33772  
 33808  
 33844  
 33880  
 33916  
 33952  
 33988  
 34024  
 34060  
 34096  
 34132  
 34168  
 34204  
 34240  
 34276  
 34312  
 34348  
 34384  
 34420  
 34456  
 34492  
 34528  
 34564  
 34600  
 34636  
 34672  
 34708  
 34744  
 34780  
 34816  
 34852  
 34888  
 34924  
 34960  
 35000  
 35036  
 35072  
 35108  
 35144  
 35180  
 35216  
 35252  
 35288  
 35324  
 35360  
 35396  
 35432  
 35468  
 35504  
 35540  
 35576  
 35612  
 35648  
 35684  
 35720  
 35756  
 35792  
 35828  
 35864  
 35900  
 35936  
 35972  
 36008  
 36044  
 36080  
 36116  
 36152  
 36188  
 36224  
 36260  
 36296  
 36332  
 36368  
 36404  
 36440  
 36476  
 36512  
 36548  
 36584  
 36620  
 36656  
 36692  
 36728  
 36764  
 36800  
 36836  
 36872  
 36908  
 36944  
 36980  
 37016  
 37052  
 37088  
 37124  
 37160  
 37196  
 37232  
 37268  
 37304  
 37340  
 37376  
 37412  
 37448  
 37484  
 37520  
 37556  
 37592  
 37628  
 37664  
 37700  
 37736  
 37772  
 37808  
 37844  
 37880  
 37916  
 37952  
 37988  
 38024  
 38060  
 38096  
 38132  
 38168  
 38204  
 38240  
 38276  
 38312  
 38348  
 38384  
 38420  
 38456  
 38492  
 38528  
 38564  
 38600  
 38636  
 38672  
 38708  
 38744  
 38780  
 38816  
 38852  
 38888  
 38924  
 38960  
 39000  
 39036  
 39072  
 39108  
 39144  
 39180  
 39216  
 39252  
 39288  
 39324  
 39360  
 39396  
 39432  
 39468  
 39504  
 39540  
 39576  
 39612  
 39648  
 39684  
 39720  
 39756  
 39792  
 39828  
 39864  
 39900  
 39936  
 39972  
 40008  
 40044  
 40080  
 40116  
 40152  
 40188  
 40224  
 40260  
 40296  
 40332  
 40368  
 40404  
 40440  
 4047



FT	DOMAIN	1747	2531	CYTOPLASMIC (POTENTIAL).		FT	DISULFID	416	429	BY SIMILARITY.
FT	DOMAIN	20	58	EGF-LIKE 1.		FT	DISULFID	423	438	BY SIMILARITY.
FT	DOMAIN	59	99	EGF-LIKE 2.		FT	DISULFID	440	449	BY SIMILARITY.
FT	DOMAIN	102	139	EGF-LIKE 3.		FT	DISULFID	456	467	BY SIMILARITY.
FT	DOMAIN	140	176	EGF-LIKE 4.		FT	DISULFID	461	476	BY SIMILARITY.
FT	DOMAIN	178	216	EGF-LIKE 5.	CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	478	487	BY SIMILARITY.
FT	DOMAIN	218	255	EGF-LIKE 6.		FT	DISULFID	494	505	BY SIMILARITY.
FT	DOMAIN	257	293	EGF-LIKE 7.	CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	499	514	BY SIMILARITY.
FT	DOMAIN	295	333	EGF-LIKE 8.	CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	516	525	BY SIMILARITY.
FT	DOMAIN	335	371	EGF-LIKE 9.	CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	532	543	BY SIMILARITY.
FT	DOMAIN	372	410	EGF-LIKE 10.		FT	DISULFID	537	552	BY SIMILARITY.
FT	DOMAIN	412	450	EGF-LIKE 11.	CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	554	563	BY SIMILARITY.
FT	DOMAIN	452	488	EGF-LIKE 12.	CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	570	580	BY SIMILARITY.
FT	DOMAIN	490	526	EGF-LIKE 13.	CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	575	589	BY SIMILARITY.
FT	DOMAIN	528	564	EGF-LIKE 14.	CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	591	600	BY SIMILARITY.
FT	DOMAIN	566	601	EGF-LIKE 15.	CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	607	618	BY SIMILARITY.
FT	DOMAIN	603	639	EGF-LIKE 16.	CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	612	627	BY SIMILARITY.
FT	DOMAIN	641	676	EGF-LIKE 17.	CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	629	638	BY SIMILARITY.
FT	DOMAIN	678	714	EGF-LIKE 18.	CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	645	655	BY SIMILARITY.
FT	DOMAIN	716	751	EGF-LIKE 19.	CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	650	664	BY SIMILARITY.
FT	DOMAIN	753	789	EGF-LIKE 20.	CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	666	675	BY SIMILARITY.
FT	DOMAIN	791	827	EGF-LIKE 21.	CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	682	693	BY SIMILARITY.
FT	DOMAIN	829	867	EGF-LIKE 22.		FT	DISULFID	687	702	BY SIMILARITY.
FT	DOMAIN	869	905	EGF-LIKE 23.	CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	704	713	BY SIMILARITY.
FT	DOMAIN	907	943	EGF-LIKE 24.		FT	DISULFID	720	730	BY SIMILARITY.
FT	DOMAIN	945	981	EGF-LIKE 25.	CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	725	739	BY SIMILARITY.
FT	DOMAIN	983	1019	EGF-LIKE 26.		FT	DISULFID	741	750	BY SIMILARITY.
FT	DOMAIN	1021	1057	EGF-LIKE 27.	CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	757	768	BY SIMILARITY.
FT	DOMAIN	1059	1095	EGF-LIKE 28.		FT	DISULFID	762	777	BY SIMILARITY.
FT	DOMAIN	1097	1143	EGF-LIKE 29.		FT	DISULFID	779	788	BY SIMILARITY.
FT	DOMAIN	1145	1181	EGF-LIKE 30.	CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	795	806	BY SIMILARITY.
FT	DOMAIN	1183	1219	EGF-LIKE 31.	CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	800	815	BY SIMILARITY.
FT	DOMAIN	1221	1265	EGF-LIKE 32.	CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	817	826	BY SIMILARITY.
FT	DOMAIN	1267	1305	EGF-LIKE 33.		FT	DISULFID	833	844	BY SIMILARITY.
FT	DOMAIN	1307	1346	EGF-LIKE 34.		FT	DISULFID	838	855	BY SIMILARITY.
FT	DOMAIN	1348	1384	EGF-LIKE 35.		FT	DISULFID	857	866	BY SIMILARITY.
FT	DOMAIN	1387	1426	EGF-LIKE 36.		FT	DISULFID	873	884	BY SIMILARITY.
FT	DOMAIN	1449	1462	CYS-RICH.		FT	DISULFID	878	893	BY SIMILARITY.
FT	REPEAT	1917	1940	ANK 1.		FT	DISULFID	895	904	BY SIMILARITY.
FT	REPEAT	1950	1986	ANK 2.		FT	DISULFID	911	922	BY SIMILARITY.
FT	REPEAT	1984	2013	ANK 3.		FT	DISULFID	916	931	BY SIMILARITY.
FT	REPEAT	2017	2046	ANK 4.		FT	DISULFID	933	942	BY SIMILARITY.
FT	REPEAT	2050	2079	ANK 5.		FT	DISULFID	987	998	BY SIMILARITY.
FT	DISULFID	24	37	BY SIMILARITY.		FT	DISULFID	992	1007	BY SIMILARITY.
FT	DISULFID	31	46	BY SIMILARITY.		FT	DISULFID	1009	1018	BY SIMILARITY.
FT	DISULFID	48	57	BY SIMILARITY.		FT	DISULFID	1025	1036	BY SIMILARITY.
FT	DISULFID	63	74	BY SIMILARITY.		FT	DISULFID	1030	1045	BY SIMILARITY.
FT	DISULFID	68	87	BY SIMILARITY.		FT	DISULFID	1047	1056	BY SIMILARITY.
FT	DISULFID	89	98	BY SIMILARITY.		FT	DISULFID	1063	1074	BY SIMILARITY.
FT	DISULFID	106	117	BY SIMILARITY.		FT	DISULFID	1068	1083	BY SIMILARITY.
FT	DISULFID	111	127	BY SIMILARITY.		FT	DISULFID	1085	1094	BY SIMILARITY.
FT	DISULFID	129	138	BY SIMILARITY.		FT	DISULFID	1101	1122	BY SIMILARITY.
FT	DISULFID	144	155	BY SIMILARITY.		FT	DISULFID	1116	1131	BY SIMILARITY.
FT	DISULFID	149	164	BY SIMILARITY.						
FT	DISULFID	166	175	BY SIMILARITY.						
FT	DISULFID	182	195	BY SIMILARITY.						
FT	DISULFID	189	204	BY SIMILARITY.						
FT	DISULFID	206	215	BY SIMILARITY.						
FT	DISULFID	222	233	BY SIMILARITY.						
FT	DISULFID	227	243	BY SIMILARITY.						
FT	DISULFID	245	254	BY SIMILARITY.						
FT	DISULFID	261	272	BY SIMILARITY.						
FT	DISULFID	266	281	BY SIMILARITY.						
FT	DISULFID	283	292	BY SIMILARITY.						
FT	DISULFID	299	312	BY SIMILARITY.						
FT	DISULFID	306	321	BY SIMILARITY.						
FT	DISULFID	323	332	BY SIMILARITY.						
FT	DISULFID	339	350	BY SIMILARITY.						
FT	DISULFID	344	359	BY SIMILARITY.						
FT	DISULFID	361	370	BY SIMILARITY.						
FT	DISULFID	376	387	BY SIMILARITY.						
FT	DISULFID	381	398	BY SIMILARITY.						
FT	DISULFID	400	409	BY SIMILARITY.						

Query Match

Best Local Similarity

Matches 115; Conservative 49; Mismatches 139; Indels 103; Gaps

QY	532	VODCPRN-CHGNCEVSGV----	CHCFPGFLGADC-----	AKAACPVLCSGN--GOYSKG	579
Db	453	VNECISNPCONDATCLDQIGEFQCICMPGYEGVYCEINTDECASSPCLHNGRCVDINEF	512		
QY	580	TCOCYSGWGAECDDVPMNOCIDPSCGGHSCIDG----	NCVSAGYKGHCE-EVDCDLP	634	
Db	513	LCQCPKGFSGHLQCYVDDECASPPCKNGAKCLDGPNTYTCVTEGYTGTHTHCEVIDECDP	572		
QY	635	TCSSHGVCVNG----	ECLCSPGWGGLNCELARVQCPCDQSGHCTYLPDTG----	LCSCDPN	687
Db	573	DPCHIGLCKDGVATFTCLCPQGYTGHHCETNINECHSQPCRGGTCQDRDNYLLCLCLKG	632		
QY	688	WMGPDCSVEVCVSDCGTH----	GVCI-----	GGACRCEEGWTGAACDQRY-----	CH-- 730
Db	633	TTGPNCNEINL--DDCASNPDCDGTCLDKIDGYECACEPYGTSMCMNVNIDECAGSPCHNG	690		



Query Match	Best Local Similarity	Score	DB 1	Length	DB 2
FT	303	312	BY SIMILARITY.		
FT	303	330	BY SIMILARITY.		
FT	303	339	BY SIMILARITY.		
FT	303	341	BY SIMILARITY.		
FT	303	356	BY SIMILARITY.		
FT	303	361	BY SIMILARITY.		
FT	303	378	BY SIMILARITY.		
FT	303	389	BY SIMILARITY.		
FT	303	409	BY SIMILARITY.		
FT	303	418	BY SIMILARITY.		
FT	303	429	BY SIMILARITY.		
FT	303	436	BY SIMILARITY.		
FT	303	447	BY SIMILARITY.		
FT	303	458	BY SIMILARITY.		
FT	303	474	BY SIMILARITY.		
FT	303	479	BY SIMILARITY.		
FT	303	496	BY SIMILARITY.		
FT	303	512	BY SIMILARITY.		
FT	303	517	BY SIMILARITY.		
FT	303	534	BY SIMILARITY.		
FT	303	550	BY SIMILARITY.		
FT	303	555	BY SIMILARITY.		
FT	303	571	BY SIMILARITY.		
FT	303	580	BY SIMILARITY.		
FT	303	597	BY SIMILARITY.		
FT	303	592	BY SIMILARITY.		
FT	303	609	BY SIMILARITY.		
FT	303	625	BY SIMILARITY.		
FT	303	640	BY SIMILARITY.		
FT	303	646	BY SIMILARITY.		
FT	303	662	BY SIMILARITY.		
FT	303	667	BY SIMILARITY.		
FT	303	684	BY SIMILARITY.		
FT	303	700	BY SIMILARITY.		
FT	303	721	BY SIMILARITY.		
FT	303	730	BY SIMILARITY.		
FT	303	739	BY SIMILARITY.		
FT	303	744	BY SIMILARITY.		
FT	303	761	BY SIMILARITY.		
FT	303	776	BY SIMILARITY.		
FT	303	781	BY SIMILARITY.		
FT	303	799	BY SIMILARITY.		
FT	303	815	BY SIMILARITY.		
FT	303	821	BY SIMILARITY.		
FT	303	838	BY SIMILARITY.		
FT	303	854	BY SIMILARITY.		
FT	303	859	BY SIMILARITY.		
FT	303	876	BY SIMILARITY.		
FT	303	892	BY SIMILARITY.		
FT	303	897	BY SIMILARITY.		
FT	303	913	BY SIMILARITY.		
FT	303	929	BY SIMILARITY.		
FT	303	934	BY SIMILARITY.		
FT	303	951	BY SIMILARITY.		
FT	303	967	BY SIMILARITY.		
FT	303	972	BY SIMILARITY.		
FT	303	989	BY SIMILARITY.		
FT	303	1005	BY SIMILARITY.		
FT	303	1010	BY SIMILARITY.		
FT	303	1025	BY SIMILARITY.		
FT	303	1041	BY SIMILARITY.		
FT	303	1056	BY SIMILARITY.		
FT	303	1073	BY SIMILARITY.		
FT	303	1089	BY SIMILARITY.		
FT	303	1094	BY SIMILARITY.		
FT	303	1111	BY SIMILARITY.		
FT	303	1127	BY SIMILARITY.		
FT	303	1132	BY SIMILARITY.		
FT	303	1149	BY SIMILARITY.		
FT	303	1165	BY SIMILARITY.		
FT	303	1177	BY SIMILARITY.		
FT	303	1194	BY SIMILARITY.		
Query Match	3.1%	Score 450.5	DB 1	Length 2318	
Best Local Similarity	26.2%	Pred. No. 2.5e-16			









AC P21783;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Neurogenic locus notch protein homolog precursor (XOTCH protein).  
 GN XOTCH.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90385285; PubMed=2402639;  
 RA Coffman C., Harris W., Kintner C.;  
 RT "Xotch, the Xenopus homolog of Drosophila notch.",  
 RL Science 249:1438-1441(1990).  
 RN [2]  
 RP REVISIONS TO 1759-1782.  
 RA Kintner C.;  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.  
 CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.  
 CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.  
 CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M33874; AB02039.1; -;  
 DR PIR; A35844; A35844.  
 DR HSP; P00740; IEDM.  
 DR InterPro; IPR002110; ANK.  
 DR InterPro; IPR000152; Asx\_hydroxyl.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR001438; EGF-II.  
 DR InterPro; IPR000800; Notch.  
 DR Pfam; PF00023; ank; 6.  
 DR Pfam; PF00008; EGF; 36.  
 DR Pfam; PF00066; notch; 3.  
 DR PRINTS; PR00010; EGFBL00D.  
 DR PRINTS; PR01452; NOTCH.  
 DR SMART; SM00248; ANK; 5.  
 DR SMART; SM00179; EGF\_CA; 23.  
 DR SMART; SM00001; EGF-like; 11.  
 DR SMART; SM00004; NL; 2.  
 DR PROSITE; PS50086; ANK\_REPEAT; 4.  
 DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 23.  
 DR PROSITE; PS00022; EGF\_1; 34.  
 DR PROSITE; PS01186; EGF\_2; 29.  
 DR PROSITE; PS01187; EGF\_CA; 21.  
 KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;  
 KW Transmembrane; Signal; Glycoprotein.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 2524 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG.  
 FT DOMAIN 20 1728 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1729 1750 POTENTIAL.  
 FT DOMAIN 1751 2524 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 20 57 EGF-LIKE 1.  
 FT DOMAIN 58 99 EGF-LIKE 2.  
 FT DOMAIN 102 140 EGF-LIKE 3.  
 FT DOMAIN 141 177 EGF-LIKE 4.  
 FT DOMAIN 179 215 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).

FT DOMAIN 217 254  
 FT DOMAIN 256 292  
 FT DOMAIN 294 332  
 FT DOMAIN 334 370  
 FT DOMAIN 371 409  
 FT DOMAIN 411 449  
 FT DOMAIN 451 487  
 FT DOMAIN 489 525  
 FT DOMAIN 527 563  
 FT DOMAIN 565 600  
 FT DOMAIN 602 638  
 FT DOMAIN 640 675  
 FT DOMAIN 677 713  
 FT DOMAIN 715 750  
 FT DOMAIN 752 788  
 FT DOMAIN 790 826  
 FT DOMAIN 828 866  
 FT DOMAIN 868 904  
 FT DOMAIN 906 942  
 FT DOMAIN 944 980  
 FT DOMAIN 982 1018  
 FT DOMAIN 1020 1056  
 FT DOMAIN 1058 1094  
 FT DOMAIN 1096 1142  
 FT DOMAIN 1144 1180  
 FT DOMAIN 1182 1218  
 FT DOMAIN 1220 1264  
 FT DOMAIN 1266 1304  
 FT DOMAIN 1306 1346  
 FT DOMAIN 1347 1383  
 FT DOMAIN 1386 1424  
 FT REPEAT 1441 1478  
 FT REPEAT 1479 1520  
 FT REPEAT 1521 1560  
 FT REPEAT 1562 1599  
 FT REPEAT 1599 1636  
 FT REPEAT 1636 1673  
 FT REPEAT 1673 1710  
 FT REPEAT 1710 1747  
 FT REPEAT 1747 1784  
 FT REPEAT 1784 1821  
 FT REPEAT 1821 1858  
 FT REPEAT 1858 1895  
 FT REPEAT 1895 1932  
 FT REPEAT 1932 1969  
 FT REPEAT 1969 2006  
 FT REPEAT 2006 2043  
 FT REPEAT 2043 2080  
 FT REPEAT 2080 2117  
 FT REPEAT 2117 2154  
 FT REPEAT 2154 2191  
 FT REPEAT 2191 2228  
 FT REPEAT 2228 2265  
 FT REPEAT 2265 2302  
 FT REPEAT 2302 2339  
 FT REPEAT 2339 2376  
 FT REPEAT 2376 2413  
 FT REPEAT 2413 2450  
 FT REPEAT 2450 2487  
 FT REPEAT 2487 2524  
 FT REPEAT 2524 2561  
 FT REPEAT 2561 2598  
 FT REPEAT 2598 2635  
 FT REPEAT 2635 2672  
 FT REPEAT 2672 2709  
 FT REPEAT 2709 2746  
 FT REPEAT 2746 2783  
 FT REPEAT 2783 2820  
 FT REPEAT 2820 2857  
 FT REPEAT 2857 2894  
 FT REPEAT 2894 2931  
 FT REPEAT 2931 2968  
 FT REPEAT 2968 3005  
 FT REPEAT 3005 3042  
 FT REPEAT 3042 3079  
 FT REPEAT 3079 3116  
 FT REPEAT 3116 3153  
 FT REPEAT 3153 3190  
 FT REPEAT 3190 3227  
 FT REPEAT 3227 3264  
 FT REPEAT 3264 3301  
 FT REPEAT 3301 3338  
 FT REPEAT 3338 3375  
 FT REPEAT 3375 3412  
 FT REPEAT 3412 3449  
 FT REPEAT 3449 3486  
 FT REPEAT 3486 3523  
 FT REPEAT 3523 3560  
 FT REPEAT 3560 3597  
 FT REPEAT 3597 3634  
 FT REPEAT 3634 3671  
 FT REPEAT 3671 3708  
 FT REPEAT 3708 3745  
 FT REPEAT 3745 3782  
 FT REPEAT 3782 3819  
 FT REPEAT 3819 3856  
 FT REPEAT 3856 3893  
 FT REPEAT 3893 3930  
 FT REPEAT 3930 3967  
 FT REPEAT 3967 4004  
 FT REPEAT 4004 4041  
 FT REPEAT 4041 4078  
 FT REPEAT 4078 4115  
 FT REPEAT 4115 4152  
 FT REPEAT 4152 4189  
 FT REPEAT 4189 4226  
 FT REPEAT 4226 4263  
 FT REPEAT 4263 4300  
 FT REPEAT 4300 4337  
 FT REPEAT 4337 4374  
 FT REPEAT 4374 4411  
 FT REPEAT 4411 4448  
 FT REPEAT 4448 4485  
 FT REPEAT 4485 4522  
 FT REPEAT 4522 4559  
 FT REPEAT 4559 4596  
 FT REPEAT 4596 4633  
 FT REPEAT 4633 4670  
 FT REPEAT 4670 4707  
 FT REPEAT 4707 4744  
 FT REPEAT 4744 4781  
 FT REPEAT 4781 4818  
 FT REPEAT 4818 4855  
 FT REPEAT 4855 4892  
 FT REPEAT 4892 4929  
 FT REPEAT 4929 4966  
 FT REPEAT 4966 5003  
 FT REPEAT 5003 5040  
 FT REPEAT 5040 5077  
 FT REPEAT 5077 5114  
 FT REPEAT 5114 5151  
 FT REPEAT 5151 5188  
 FT REPEAT 5188 5225  
 FT REPEAT 5225 5262  
 FT REPEAT 5262 5299  
 FT REPEAT 5299 5336  
 FT REPEAT 5336 5373  
 FT REPEAT 5373 5410  
 FT REPEAT 5410 5447  
 FT REPEAT 5447 5484  
 FT REPEAT 5484 5521  
 FT REPEAT 5521 5558  
 FT REPEAT 5558 5595  
 FT REPEAT 5595 5632  
 FT REPEAT 5632 5669  
 FT REPEAT 5669 5706  
 FT REPEAT 5706 5743  
 FT REPEAT 5743 5780  
 FT REPEAT 5780 5817  
 FT REPEAT 5817 5854  
 FT REPEAT 5854 5891  
 FT REPEAT 5891 5928  
 FT REPEAT 5928 5965  
 FT REPEAT 5965 6002  
 FT REPEAT 6002 6039  
 FT REPEAT 6039 6076  
 FT REPEAT 6076 6113  
 FT REPEAT 6113 6150  
 FT REPEAT 6150 6187  
 FT REPEAT 6187 6224  
 FT REPEAT 6224 6261  
 FT REPEAT 6261 6298  
 FT REPEAT 6298 6335  
 FT REPEAT 6335 6372  
 FT REPEAT 6372 6409  
 FT REPEAT 6409 6446  
 FT REPEAT 6446 6483  
 FT REPEAT 6483 6520  
 FT REPEAT 6520 6557  
 FT REPEAT 6557 6594  
 FT REPEAT 6594 6631  
 FT REPEAT 6631 6668  
 FT REPEAT 6668 6705  
 FT REPEAT 6705 6742  
 FT REPEAT 6742 6779  
 FT REPEAT 6779 6816  
 FT REPEAT 6816 6853  
 FT REPEAT 6853 6890  
 FT REPEAT 6890 6927  
 FT REPEAT 6927 6964  
 FT REPEAT 6964 7001  
 FT REPEAT 7001 7038  
 FT REPEAT 7038 7075  
 FT REPEAT 7075 7112  
 FT REPEAT 7112 7149  
 FT REPEAT 7149 7186  
 FT REPEAT 7186 7223  
 FT REPEAT 7223 7260  
 FT REPEAT 7260 7297  
 FT REPEAT 7297 7334  
 FT REPEAT 7334 7371  
 FT REPEAT 7371 7408  
 FT REPEAT 7408 7445  
 FT REPEAT 7445 7482  
 FT REPEAT 7482 7519  
 FT REPEAT 7519 7556  
 FT REPEAT 7556 7593  
 FT REPEAT 7593 7630  
 FT REPEAT 7630 7667  
 FT REPEAT 7667 7704  
 FT REPEAT 7704 7741  
 FT REPEAT 7741 7778  
 FT REPEAT 7778 7815  
 FT REPEAT 7815 7852  
 FT REPEAT 7852 7889  
 FT REPEAT 7889 7926  
 FT REPEAT 7926 7963  
 FT REPEAT 7963 8000  
 FT REPEAT 8000 8037  
 FT REPEAT 8037 8074  
 FT REPEAT 8074 8111  
 FT REPEAT 8111 8148  
 FT REPEAT 8148 8185  
 FT REPEAT 8185 8222  
 FT REPEAT 8222 8259  
 FT REPEAT 8259 8296  
 FT REPEAT 8296 8333  
 FT REPEAT 8333 8370  
 FT REPEAT 8370 8407  
 FT REPEAT 8407 8444  
 FT REPEAT 8444 8481  
 FT REPEAT 8481 8518  
 FT REPEAT 8518 8555  
 FT REPEAT 8555 8592  
 FT REPEAT 8592 8629  
 FT REPEAT 8629 8666  
 FT REPEAT 8666 8703  
 FT REPEAT 8703 8740  
 FT REPEAT 8740 8777  
 FT REPEAT 8777 8814  
 FT REPEAT 8814 8851  
 FT REPEAT 8851 8888  
 FT REPEAT 8888 8925  
 FT REPEAT 8925 8962  
 FT REPEAT 8962 9000  
 FT REPEAT 9000 9037  
 FT REPEAT 9037 9074  
 FT REPEAT 9074 9111  
 FT REPEAT 9111 9148  
 FT REPEAT 9148 9185  
 FT REPEAT 9185 9222  
 FT REPEAT 9222 9259  
 FT REPEAT 9259 9296  
 FT REPEAT 9296 9333  
 FT REPEAT 9333 9370  
 FT REPEAT 9370 9407  
 FT REPEAT 9407 9444  
 FT REPEAT 9444 9481  
 FT REPEAT 9481 9518  
 FT REPEAT 9518 9555  
 FT REPEAT 9555 9592  
 FT REPEAT 9592 9629  
 FT REPEAT 9629 9666  
 FT REPEAT 9666 9703  
 FT REPEAT 9703 9740  
 FT REPEAT 9740 9777  
 FT REPEAT 9777 9814  
 FT REPEAT 9814 9851  
 FT REPEAT 9851 9888  
 FT REPEAT 9888 9925  
 FT REPEAT 9925 9962  
 FT REPEAT 9962 10000  
 FT REPEAT 10000 10037  
 FT REPEAT 10037 10074  
 FT REPEAT 10074 10111  
 FT REPEAT 10111 10148  
 FT REPEAT 10148 10185  
 FT REPEAT 10185 10222  
 FT REPEAT 10222 10259  
 FT REPEAT 10259 10296  
 FT REPEAT 10296 10333  
 FT REPEAT 10333 10370  
 FT REPEAT 10370 10407  
 FT REPEAT 10407 10444  
 FT REPEAT 10444 10481  
 FT REPEAT 10481 10518  
 FT REPEAT 10518 10555  
 FT REPEAT 10555 10592  
 FT REPEAT 10592 10629  
 FT REPEAT 10629 10666  
 FT REPEAT 10666 10703  
 FT REPEAT 10703 10740  
 FT REPEAT 10740 10777  
 FT REPEAT 10777 10814  
 FT REPEAT 10814 10851  
 FT REPEAT 10851 10888  
 FT REPEAT 10888 10925  
 FT REPEAT 10925 10962  
 FT REPEAT 10962 11000  
 FT REPEAT 11000 11037  
 FT REPEAT 11037 11074  
 FT REPEAT 11074 11111  
 FT REPEAT 11111 11148  
 FT REPEAT 11148 11185  
 FT REPEAT 11185 11222  
 FT REPEAT 11222 11259  
 FT REPEAT 11259 11296  
 FT REPEAT 11296 11333  
 FT REPEAT 11333 11370  
 FT REPEAT 11370 11407  
 FT REPEAT 11407 11444  
 FT REPEAT 11444 11481  
 FT REPEAT 11481 11518  
 FT REPEAT 11518 11555  
 FT REPEAT 11555 11592  
 FT REPEAT 11592 11629  
 FT REPEAT 11629 11666  
 FT REPEAT 11666 11703  
 FT REPEAT 11703 11740  
 FT REPEAT 11740 11777  
 FT REPEAT 11777 11814  
 FT REPEAT 11814 11851  
 FT REPEAT 11851 11888  
 FT REPEAT 11888 11925  
 FT REPEAT 11925 11962  
 FT REPEAT 11962 12000  
 FT REPEAT 12000 12037  
 FT REPEAT 12037 12074  
 FT REPEAT 12074 12111  
 FT REPEAT 12111 12148  
 FT REPEAT 12148 12185  
 FT REPEAT 12185 12222  
 FT REPEAT 12222 12259  
 FT REPEAT 12259 12296  
 FT REPEAT 12296 12333  
 FT REPEAT 12333 12370  
 FT REPEAT 12370 12407  
 FT REPEAT 12407 12444  
 FT REPEAT 12444 12481  
 FT REPEAT 12481 12518  
 FT REPEAT 12518 12555  
 FT REPEAT 12555 12592  
 FT REPEAT 12592 12629  
 FT REPEAT 12629 12666  
 FT REPEAT 12666 12703  
 FT REPEAT 12703 12740  
 FT REPEAT 12740 12777  
 FT REPEAT 12777 12814  
 FT REPEAT 12814 12851  
 FT REPEAT 12851 12888  
 FT REPEAT 12888 12925  
 FT REPEAT 12925 12962  
 FT REPEAT 12962 13000  
 FT REPEAT 13000 13037  
 FT REPEAT 13037 13074  
 FT REPEAT 13074 13111  
 FT REPEAT 13111 13148  
 FT REPEAT 13148 13185  
 FT REPEAT 13185 13222  
 FT REPEAT 13222 13259  
 FT REPEAT 13259 13296  
 FT REPEAT 13296 13333  
 FT REPEAT 13333 13370  
 FT REPEAT 13370 13407  
 FT REPEAT 13407 13444  
 FT REPEAT 13444 13481  
 FT REPEAT 13481 13518  
 FT REPEAT 13518 13555  
 FT REPEAT 13555 13592  
 FT REPEAT 13592 13629  
 FT REPEAT 13629 13666  
 FT REPEAT 13666 13703  
 FT REPEAT 13703 13740  
 FT REPEAT 13740 13777  
 FT REPEAT 13777 13814  
 FT REPEAT 13814 13851  
 FT REPEAT 13851 13888  
 FT REPEAT 13888 13925  
 FT REPEAT 13925 13962  
 FT REPEAT 13962 14000  
 FT REPEAT 14000 14037  
 FT REPEAT 14037 14074  
 FT REPEAT 14074 14111  
 FT REPEAT 14111 14148  
 FT REPEAT 14148 14185  
 FT REPEAT 14185 14222  
 FT REPEAT 14222 14259  
 FT REPEAT 14259 14296  
 FT REPEAT 14296 14333  
 FT REPEAT 14333 14370  
 FT REPEAT 14370 14407  
 FT REPEAT 14407 14444  
 FT REPEAT 14444 14481  
 FT REPEAT 14481 14518  
 FT REPEAT 14518 14555  
 FT REPEAT 14555 14592  
 FT REPEAT 14592 14629  
 FT REPEAT 14629 14666  
 FT REPEAT 14666 14703  
 FT REPEAT 14703 14740  
 FT REPEAT 14740 14777  
 FT REPEAT 14777 14814  
 FT REPEAT 14814 14851  
 FT REPEAT 14851 14888  
 FT REPEAT 14888 14925  
 FT REPEAT 14925 14962  
 FT REPEAT 14962 15000  
 FT REPEAT 15000 15037  
 FT REPEAT 15037 15074  
 FT REPEAT 15074 15111  
 FT REPEAT 15111 15148  
 FT REPEAT 15148 15185  
 FT REPEAT 15185 15222  
 FT REPEAT 15222 15259  
 FT REPEAT 15259 15296  
 FT REPEAT 15296 15333  
 FT REPEAT 15333 15370  
 FT REPEAT 15370 15407  
 FT REPEAT 15407 15444  
 FT REPEAT 15444 15481  
 FT REPEAT 15481 15518  
 FT REPEAT 15518 15555  
 FT REPEAT 15555 15592  
 FT REPEAT 15592 15629  
 FT REPEAT 15629 15666  
 FT REPEAT 15666 15703  
 FT REPEAT 15703 15740  
 FT REPEAT 15740 15777  
 FT REPEAT 15777 15814  
 FT REPEAT 15814 15851  
 FT REPEAT 15851 15888  
 FT REPEAT 15888 15925  
 FT REPEAT 15925 15962  
 FT REPEAT 15962 16000  
 FT REPEAT 16000 16037  
 FT REPEAT 16037 16074  
 FT REPEAT 16074 16111  
 FT REPEAT 16111 16148  
 FT REPEAT 16148 16185  
 FT REPEAT 16185 16222  
 FT REPEAT 16222 16259  
 FT REPEAT 16259 16296  
 FT REPEAT 16296 16333  
 FT REPEAT 16333 16370  
 FT REPEAT 16370 16407  
 FT REPEAT 16407 16444  
 FT REPEAT 16444 16481  
 FT REPEAT 16481 16518  
 FT REPEAT 16518 16555  
 FT REPEAT 16555 16592  
 FT REPEAT 16592 16629  
 FT REPEAT 16629 16666  
 FT REPEAT 16666 16703  
 FT REPEAT 16703 16740  
 FT REPEAT 16740 16777  
 FT REPEAT 16777 16814  
 FT REPEAT 16814 16851  
 FT REPEAT 16851 16888  
 FT REPEAT 16888 16925  
 FT REPEAT 16925 16962  
 FT REPEAT 16962 17000  
 FT REPEAT 17000 17037  
 FT REPEAT 17037 17074  
 FT REPEAT 17074 17111  
 FT REPEAT 17111 17148  
 FT REPEAT 17148 17185  
 FT REPEAT 17185 17222  
 FT REPEAT 17222 17259  
 FT REPEAT 17259 17296  
 FT REPEAT 17296 17333  
 FT REPEAT 17333 17370  
 FT REPEAT 17370 17407  
 FT REPEAT 17407 17444  
 FT REPEAT 17444 17481  
 FT REPEAT 17481 17518  
 FT REPEAT 17518 17555  
 FT REPEAT 17555 17592  
 FT REPEAT 17592 17629  
 FT REPEAT 17629 17666  
 FT REPEAT 17666 17703  
 FT REPEAT 17703 17740  
 FT REPEAT 17740 17777  
 FT REPEAT 17777 17814  
 FT REPEAT 17814 17851  
 FT REPEAT 17851 17888  
 FT REPEAT 17888 17925  
 FT REPEAT 17925 17962  
 FT REPEAT 17962 18000  
 FT REPEAT 18000 18037  
 FT REPEAT 18037 18074  
 FT REPEAT 18074 18111  
 FT REPEAT 18111 18148  
 FT REPEAT 18148 18185  
 FT REPEAT 18185 18222  
 FT REPEAT 18222 18259  
 FT REPEAT 18259 18296  
 FT REPEAT 18296 18333  
 FT REPEAT 18333 18370  
 FT REPEAT 18370 18407  
 FT REPEAT 18407 18444  
 FT REPEAT 18444 18481  
 FT REPEAT 18481 18518  
 FT REPEAT 18518 18555  
 FT REPEAT 18555 18592  
 FT REPEAT 18592 18629  
 FT REPEAT 18629 18666  
 FT REPEAT 18666 18703  
 FT REPEAT 18703 18740  
 FT REPEAT 18740 18777  
 FT REPEAT 18777 18814  
 FT REPEAT 18814 18851  
 FT REPEAT 18851 18888  
 FT REPEAT 18888 18925  
 FT REPEAT 18925 18962  
 FT REPEAT 18962 19000  
 FT REPEAT 19000 19037  
 FT REPEAT 19037 19074  
 FT REPEAT 19074 19111  
 FT REPEAT 19111 19148  
 FT REPEAT 19148 19185  
 FT REPEAT 19185 19222  
 FT REPEAT 19222 19259  
 FT REPEAT 19259 19296  
 FT REPEAT 19296 19333  
 FT REPEAT 19333 19370  
 FT REPEAT 19370 19407  
 FT REPEAT 19407 19444  
 FT REPEAT 19444 19481  
 FT REPEAT 19481 19518  
 FT REPEAT 19518 19555  
 FT REPEAT 19555 19592  
 FT REPEAT 19592 19629  
 FT REPEAT 19629 19666  
 FT REPEAT 19666 19703  
 FT REPEAT 19703 19740  
 FT REPEAT 19740 19777  
 FT REPEAT 19777 19814  
 FT REPEAT 19814 19851  
 FT REPEAT 19851 19888  
 FT REPEAT 19888 19925  
 FT REPEAT 19925 19962  
 FT REPEAT 19962 20000  
 FT REPEAT 20000 20037  
 FT REPEAT 20037 20074  
 FT REPEAT 20074 20111  
 FT REPEAT 20111 20148  
 FT REPEAT 20148 20185  
 FT REPEAT 20185 20222  
 FT REPEAT 20222 20259  
 FT REPEAT 20259 20296  
 FT REPEAT 20296 20333  
 FT REPEAT 20333 20370  
 FT REPEAT 20370 20407  
 FT REPEAT 20407 20444  
 FT REPEAT 20444 20481  
 FT REPEAT 20481 20518  
 FT REPEAT 20518 20555  
 FT REPEAT 20555 20592  
 FT REPEAT 20592 20629  
 FT REPEAT 20629 20666  
 FT REPEAT 20666 20703  
 FT REPEAT 20703 20740  
 FT REPEAT 20740 20777  
 FT REPEAT 20777 20814  
 FT REPEAT 20814 20851  
 FT REPEAT 20851 20888  
 FT REPEAT 20888 20925  
 FT REPEAT 20925 20962  
 FT REPEAT 20962 21000  
 FT REPEAT 21000 21037  
 FT REPEAT 21037 21074  
 FT REPEAT 21074 21111  
 FT REPEAT 21111 21148  
 FT REPEAT 21148 21185  
 FT REPEAT 21185 21222  
 FT REPEAT 21222 21259  
 FT REPEAT 21259 21296  
 FT REPEAT 21296 21333  
 FT REPEAT 21333 21370  
 FT REPEAT 21370 21407  
 FT REPEAT 21407 21444  
 FT REPEAT 21444 21481  
 FT REPEAT 21481 21518  
 FT REPEAT 21518 21555  
 FT REPEAT 21555 21592  
 FT REPEAT 21592 21629  
 FT REPEAT 21629 21666  
 FT REPEAT 21666 21703  
 FT REPEAT 21703 21740  
 FT REPEAT 21740 21777  
 FT REPEAT 21777 21814  
 FT REPEAT 21814 21851  
 FT REPEAT 21851 21888  
 FT REPEAT 21888 21925  
 FT REPEAT 21925 21962  
 FT REPEAT 21962 22000  
 FT REPEAT 22000 22037  
 FT REPEAT 22037 22074  
 FT REPEAT 22074 22111  
 FT REPEAT 22111 22148  
 FT REPEAT 22148 22185  
 FT REPEAT 22185 22222

Query Match 2.98; Score 428.5; DB 1; Length 2524;  
 Best Local Similarity 28.1%; Pred. No. 4.7e-15;  
 Matches 123; Conservative 42; Mismatches 137; Indels 135; Gaps 26;

Y 532 VQDC-PRNCHGECVSGV-----CHCPFLGADCAACAPVL---CSGNGQYS----- 577  
 DB 566 INECIPDPCH-YGCKDGIATFTCLCRPGYGRCLDNDINECLSKPCLNGGQCTDRENGY 624  
 Y 578 -----KGT-----CQYSGWKGAECDPVNMQCIDP 602  
 DB 625 ICTCPKGTGVNCEKTKIDDCASNLCDNGKCIDKIDGVECTCEPGYTGKLCININECDN 684  
 Y 603 SCGGHGSC---IDG-NCVCSAGYKGEHC--EEVCLDPTCSSHGVC---VNG-BCLCSGP 652  
 DB 685 PCRNGGTCKDQINGFTCVCPDGYHDHMCLESEVNECNSNPC-IHGACHDGVNGYKDCDEAG 743  
 Y 653 WGLNCELARVQC-PDQCSGHGTPLPTDG--LCSDPMMWGPDCSVEV--CSVD-CGTHG 706  
 DB 744 WGSNCNDINNCEBSPCMWNGGTCKDMTGAYICTCKAGFGSPNCOTNINECSNPNLNHG 803

FT DISULFID 455 466 BY SIMILARITY.  
 FT DISULFID 460 475 BY SIMILARITY.  
 FT DISULFID 477 486 BY SIMILARITY.  
 FT DISULFID 493 504 BY SIMILARITY.  
 FT DISULFID 498 513 BY SIMILARITY.  
 FT DISULFID 515 524 BY SIMILARITY.  
 FT DISULFID 531 542 BY SIMILARITY.  
 FT DISULFID 536 551 BY SIMILARITY.  
 FT DISULFID 553 562 BY SIMILARITY.  
 FT DISULFID 569 579 BY SIMILARITY.  
 FT DISULFID 574 588 BY SIMILARITY.  
 FT DISULFID 590 599 BY SIMILARITY.  
 FT DISULFID 606 617 BY SIMILARITY.  
 FT DISULFID 611 626 BY SIMILARITY.  
 FT DISULFID 628 637 BY SIMILARITY.  
 FT DISULFID 644 654 BY SIMILARITY.  
 FT DISULFID 649 663 BY SIMILARITY.  
 FT DISULFID 665 674 BY SIMILARITY.  
 FT DISULFID 681 692 BY SIMILARITY.  
 FT DISULFID 686 701 BY SIMILARITY.  
 FT DISULFID 703 712 BY SIMILARITY.  
 FT DISULFID 719 729 BY SIMILARITY.  
 FT DISULFID 724 738 BY SIMILARITY.  
 FT DISULFID 740 749 BY SIMILARITY.  
 FT DISULFID 756 767 BY SIMILARITY.  
 FT DISULFID 761 776 BY SIMILARITY.  
 FT DISULFID 778 787 BY SIMILARITY.  
 FT DISULFID 794 805 BY SIMILARITY.  
 FT DISULFID 799 814 BY SIMILARITY.  
 FT DISULFID 816 825 BY SIMILARITY.  
 FT DISULFID 832 843 BY SIMILARITY.  
 FT DISULFID 837 854 BY SIMILARITY.  
 FT DISULFID 856 865 BY SIMILARITY.  
 FT DISULFID 872 883 BY SIMILARITY.  
 FT DISULFID 877 892 BY SIMILARITY.  
 FT DISULFID 894 903 BY SIMILARITY.  
 FT DISULFID 910 921 BY SIMILARITY.  
 FT DISULFID 915 930 BY SIMILARITY.  
 FT DISULFID 932 941 BY SIMILARITY.  
 FT DISULFID 986 997 BY SIMILARITY.  
 FT DISULFID 991 1006 BY SIMILARITY.  
 FT DISULFID 1008 1017 BY SIMILARITY.  
 FT DISULFID 1024 1035 BY SIMILARITY.  
 FT DISULFID 1029 1044 BY SIMILARITY.  
 FT DISULFID 1046 1055 BY SIMILARITY.  
 FT DISULFID 1062 1073 BY SIMILARITY.  
 FT DISULFID 1067 1082 BY SIMILARITY.  
 FT DISULFID 1084 1093 BY SIMILARITY.  
 FT DISULFID 1100 1121 BY SIMILARITY.  
 FT DISULFID 1115 1130 BY SIMILARITY.  
 FT DISULFID 1132 1141 BY SIMILARITY.  
 FT DISULFID 1148 1159 BY SIMILARITY.  
 FT DISULFID 1153 1168 BY SIMILARITY.

707 VCI----GGACRCBEGWTGAACDQVCHP-----RCIEHGTCKDKG-----CECREGNG 752  
 DB 804 TCIDDVAGYKNCMLPTYGAI-BAVLAPCAGSPCKNGGRCKESEDFTFSCBPPGHWG 862  
 Y 753 EHTIG-----ROTAGT-----ETDGC-PDLCHNGRGT 780  
 DB 863 QTCIDMNECVNRPCRGATCQNTNGSKCKPGYTRNCMDIDDCQPNCHNGSGS 922  
 Y 781 LGONSWQVCOTGARGPCCNVAMETSCADNKDNEGDGLVDCLD----- 823  
 DB 923 DGINFFCNCBPAGRPKCEEDI-NECASNPCKNGANTDCVNSYTCCTQCPFGSGIHES 981  
 Y 824 --PDCCQLQSACQNSLLC 838  
 DB 982 NTPD-CTESSCFNGGTC 997

RESULT 11  
 CRB\_DROME  
 ID CRB\_DROME STANDARD; PRT; 2139 AA.  
 AC P10040:  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Crumbs protein precursor (95F).  
 GN CRB.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-OREGON-R; TISSUE=Embryo;  
 RA MEDLINE-90263104; PubMed-2344615;  
 RA Tepass U., Theres C., Knust E.;  
 RT "Crumbs encodes an EGF-like protein expressed on apical membranes of  
 epithelia.";  
 RL Cell 61:787-799(1990).  
 RN [2]  
 RP SEQUENCE OF 1663-1955 FROM N.A.  
 RC TISSUE=Embryo;  
 RA MEDLINE-87218537; PubMed-3107986;  
 RA Knust E., Dietrich U., Tepass U., Bremer K.A., Weigel D.,  
 RA Vaessin H., Campos-Ortega J.A.;  
 RT "EGF homologous sequences encoded in the genome of Drosophila  
 melanogaster, and their relation to neurogenic genes.";  
 RL EMBO J. 6:761-766(1987).  
 CC -|- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,  
 CC POSSIBLY FOR THE ESTABLISHMENT AND/OR MAINTENANCE OF CELL  
 CC POLARITY. IT MAY ACT AS A SIGNAL.  
 CC -|- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -|- PM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).  
 CC -|- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.  
 CC -|- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M33753; AAA28428.1; ALT\_SEQ.  
 DR EMBL; X05144; CAA28793.1; -  
 DR PIR; B26637; B26637.  
 DR PIR; A35672; A35672.  
 DR HSP; P00740; IIXA.  
 DR FlyBase; FBgn0000368; crb.  
 DR InterPro; IPR000152; Asx\_hydroxyl.

DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR000742; EGF\_2.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR001438; EGF\_II.  
 DR InterPro: IPR001791; Laminin\_G.  
 DR Pfam: PF00008; EGF; 26.  
 DR Pfam: PF00054; laminin\_G; 3.  
 DR PRINTS: PRO0010; EGFBL00D.  
 DR SMART: SM00179; EGF\_Ca; 11.  
 DR SMART: SM00001; EGF\_Like; 16.  
 DR SMART: SM00282; LamG; 3.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 15.  
 DR PROSITE: PS00022; EGF\_1; 26.  
 DR PROSITE: PS01186; EGF\_2; 17.  
 DR PROSITE: PS01187; EGF\_Ca; 15.  
 DR PROSITE: PS50025; LAM\_G\_DOMAIN; 3.  
 KW Differentiation; Repeat; EGF-like domain; Transmembrane;  
 KW Glycoprotein; Signal; Phosphorylation.  
 FT SIGNAL 1 90  
 FT CHAIN 91 2139 CRUMBS PROTEIN.  
 FT DOMAIN 91 2084 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 2085 2111 POTENTIAL.  
 FT DOMAIN 2112 2139 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 267 303 EGF-LIKE 1.  
 FT DOMAIN 306 343 EGF-LIKE 2.  
 FT DOMAIN 348 386 EGF-LIKE 3.  
 FT DOMAIN 388 425 EGF-LIKE 4.  
 FT DOMAIN 427 463 EGF-LIKE 5.  
 FT DOMAIN 464 500 EGF-LIKE 6.  
 FT DOMAIN 501 532 EGF-LIKE 7.  
 FT DOMAIN 545 581 EGF-LIKE 8.  
 FT DOMAIN 582 611 EGF-LIKE 9.  
 FT DOMAIN 609 646 EGF-LIKE 10.  
 FT DOMAIN 648 685 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 687 723 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 725 761 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 763 800 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 802 838 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 840 902 EGF-LIKE 16.  
 FT DOMAIN 904 940 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 942 978 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 980 1021 EGF-LIKE 19.  
 FT DOMAIN 1023 1205 LAMININ G-LIKE 1.  
 FT DOMAIN 1207 1243 EGF-LIKE 20.  
 FT DOMAIN 1250 1480 LAMININ G-LIKE 2.  
 FT DOMAIN 1481 1517 EGF-LIKE 21.  
 FT DOMAIN 1558 1758 LAMININ G-LIKE 3.  
 FT DOMAIN 1759 1795 EGF-LIKE 22.  
 FT DOMAIN 1797 1833 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 1835 1871 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 1874 1915 EGF-LIKE 25.  
 FT DOMAIN 1915 1951 EGF-LIKE 26.  
 FT DOMAIN 1953 1989 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 1991 2029 EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 2030 2070 EGF-LIKE 29.  
 FT DISULFID 271 282 BY SIMILARITY.  
 FT DISULFID 276 291 BY SIMILARITY.  
 FT DISULFID 293 302 BY SIMILARITY.  
 FT DISULFID 310 321 BY SIMILARITY.  
 FT DISULFID 315 331 BY SIMILARITY.  
 FT DISULFID 333 342 BY SIMILARITY.  
 FT DISULFID 352 363 BY SIMILARITY.  
 FT DISULFID 357 374 BY SIMILARITY.  
 FT DISULFID 376 385 BY SIMILARITY.  
 FT DISULFID 392 403 BY SIMILARITY.  
 FT DISULFID 397 412 BY SIMILARITY.  
 FT DISULFID 414 424 BY SIMILARITY.  
 FT DISULFID 431 442 BY SIMILARITY.  
 FT DISULFID 436 451 BY SIMILARITY.  
 FT DISULFID 453 462 BY SIMILARITY.  
 FT DISULFID 468 479 BY SIMILARITY.  
 FT DISULFID 473 488 BY SIMILARITY.  
 FT DISULFID 490 499 BY SIMILARITY.  
 FT DISULFID 505 515 BY SIMILARITY.  
 FT DISULFID 509 515 BY SIMILARITY.  
 FT DISULFID 522 531 BY SIMILARITY.  
 FT DISULFID 549 562 BY SIMILARITY.  
 FT DISULFID 556 569 BY SIMILARITY.  
 FT DISULFID 571 580 BY SIMILARITY.  
 FT DISULFID 586 597 BY SIMILARITY.  
 FT DISULFID 591 602 BY SIMILARITY.  
 FT DISULFID 604 610 BY SIMILARITY.  
 FT DISULFID 613 624 BY SIMILARITY.  
 FT DISULFID 618 634 BY SIMILARITY.  
 FT DISULFID 636 645 BY SIMILARITY.  
 FT DISULFID 652 664 BY SIMILARITY.  
 FT DISULFID 659 673 BY SIMILARITY.  
 FT DISULFID 675 684 BY SIMILARITY.  
 FT DISULFID 691 702 BY SIMILARITY.  
 FT DISULFID 696 711 BY SIMILARITY.  
 FT DISULFID 713 722 BY SIMILARITY.  
 FT DISULFID 729 740 BY SIMILARITY.  
 FT DISULFID 734 749 BY SIMILARITY.  
 FT DISULFID 751 760 BY SIMILARITY.  
 FT DISULFID 767 778 BY SIMILARITY.  
 FT DISULFID 772 787 BY SIMILARITY.  
 FT DISULFID 789 799 BY SIMILARITY.  
 FT DISULFID 806 817 BY SIMILARITY.  
 FT DISULFID 811 826 BY SIMILARITY.  
 FT DISULFID 828 837 BY SIMILARITY.  
 FT DISULFID 844 855 BY SIMILARITY.  
 FT DISULFID 849 890 BY SIMILARITY.  
 FT DISULFID 892 901 BY SIMILARITY.  
 FT DISULFID 908 919 BY SIMILARITY.  
 FT DISULFID 913 928 BY SIMILARITY.  
 FT DISULFID 930 939 BY SIMILARITY.  
 FT DISULFID 946 957 BY SIMILARITY.  
 FT DISULFID 952 966 BY SIMILARITY.  
 FT DISULFID 968 977 BY SIMILARITY.  
 FT DISULFID 984 995 BY SIMILARITY.  
 FT DISULFID 989 1009 BY SIMILARITY.  
 FT DISULFID 1011 1020 BY SIMILARITY.  
 FT DISULFID 1211 1222 BY SIMILARITY.  
 FT DISULFID 1216 1231 BY SIMILARITY.  
 FT DISULFID 1233 1242 BY SIMILARITY.  
 FT DISULFID 1485 1496 BY SIMILARITY.  
 FT DISULFID 1490 1505 BY SIMILARITY.  
 FT DISULFID 1507 1516 BY SIMILARITY.  
 FT DISULFID 1763 1774 BY SIMILARITY.  
 FT DISULFID 1768 1783 BY SIMILARITY.  
 FT DISULFID 1785 1794 BY SIMILARITY.  
 FT DISULFID 1801 1812 BY SIMILARITY.  
 FT DISULFID 1806 1821 BY SIMILARITY.  
 FT DISULFID 1823 1832 BY SIMILARITY.  
 FT DISULFID 1839 1850 BY SIMILARITY.  
 FT DISULFID 1844 1859 BY SIMILARITY.  
 FT DISULFID 1861 1870 BY SIMILARITY.  
 FT DISULFID 1878 1889 BY SIMILARITY.  
 FT DISULFID 1883 1903 BY SIMILARITY.  
 FT DISULFID 1905 1914 BY SIMILARITY.  
 FT DISULFID 1919 1930 BY SIMILARITY.  
 FT DISULFID 1924 1939 BY SIMILARITY.  
 FT DISULFID 1941 1950 BY SIMILARITY.  
 FT DISULFID 1957 1968 BY SIMILARITY.  
 FT DISULFID 1962 1977 BY SIMILARITY.  
 FT DISULFID 1979 1988 BY SIMILARITY.  
 FT DISULFID 1995 2008 BY SIMILARITY.  
 FT DISULFID 2002 2017 BY SIMILARITY.  
 FT DISULFID 2019 2028 BY SIMILARITY.  
 FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 400 400 N-LINKED (GLCNAC. . .) (POTENTIAL).





RA Adachi H., Tsujimoto M., Arai H., Inoue K.;  
 RT "Expression cloning of a novel scavenger receptor from human  
 RT endothelial cells."  
 RL J. Biol. Chem. 272:31217-31220(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BONE marrow;  
 RX MEDLINE=96127530; PubMed=8590280;  
 RA Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;  
 RT "Prediction of the coding sequences of unidentified human genes. IV.  
 RT The coding sequences of 40 new genes (K1AA0121-K1AA0160) deduced by  
 RT analysis of cDNA clones from human cell line KG-1."  
 RL DNA Res. 2:167-174(1995).  
 CC -1- FUNCTION: Mediates the binding and degradation of acetylated low  
 CC density lipoprotein (Ac-LDL).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).  
 CC -1- TISSUE SPECIFICITY: Endothelial cells.  
 CC -1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; D86864; BAA24070.1; -;  
 DR EMBL; D63483; BAA09770.1; -;  
 DR HSPF; P01180; 2BN2.  
 DR InterPro; IPR000561; EGF-like.  
 DR SMART; SM00181; EGF; 5.  
 DR SMART; SM00001; EGF-like; 2.  
 DR PROSITE; PS00022; EGF\_1; 6.  
 DR PROSITE; PS01186; EGF\_2; 6.  
 KW Receptor; Glycoprotein; Signal; Transmembrane; EGF-like domain;  
 KW Repeat.  
 FT SIGNAL 1 19  
 FT CHAIN 20 830  
 FT DOMAIN 20 421  
 FT TRANSMEM 422 442  
 FT DOMAIN 443 830  
 FT DOMAIN 53 87  
 FT DOMAIN 95 130  
 FT DOMAIN 155 191  
 FT DOMAIN 215 249  
 FT DOMAIN 302 339  
 FT DOMAIN 351 382  
 FT DOMAIN 476 620  
 FT DOMAIN 622 798  
 FT DOMAIN 431 438  
 FT DISULFID 57 69  
 FT DISULFID 63 75  
 FT DISULFID 77 86  
 FT DISULFID 99 111  
 FT DISULFID 105 118  
 FT DISULFID 120 129  
 FT DISULFID 159 172  
 FT DISULFID 165 179  
 FT DISULFID 181 190  
 FT DISULFID 219 230  
 FT DISULFID 225 237  
 FT DISULFID 239 248  
 FT DISULFID 306 319  
 FT DISULFID 313 326  
 FT DISULFID 329 338  
 FT DISULFID 355 363  
 FT DISULFID 358 370  
 FT DISULFID 372 381  
 FT CARBOHYD 289 289  
 FT CARBOHYD 382 382  
 FT CARBOHYD 393 393  
 FT CONFLICT 662 662

SQ SEQUENCE 830 AA; 87430 MW; F560D9E1AA64D779 CRC64;  
 Query Match 2.8%; Score 415.5; DB 1; Length 830;  
 Best Local Similarity 25.5%; Pred. No. 4.le-15;  
 Matches 138; Conservative 49; Mismatches 189; Indels 165; Gaps 28;  
 Qy 533 QDC-----PRNCHNGGECYS-GVCHCFPGFLGADCA-----KAACPVLCSG 572  
 Db 50 QECITPICEGPDACOKDEVKVGKGLCRCKPGFFGAHCSRRCPGQVWGPDRCRSCPHPHG 109  
 Qy 573 NGQYSKGTCCQYSGWKGAEDVPMNQCIDPSCGHGSC--IDGNCVCSAGYKGEHVD 630  
 Db 110 QCEPATGACQCAQADRWGARGCEFP---C---ACGPHGRCDPATGCHCEPGWMSSTCR-P 162  
 Qy 631 CLDPTCSSHGVGVNGECLSPGWGLNCELARVQCPCDSGHGTYL-PDTGLCSGCDPMNM 689  
 Db 163 CQNTAAARCEQATGACVCKPQWGRRCSP-----RCNCHGSPCEDSGRCACRPGMW 215  
 Qy 690 GPDGSVEV-----CSVDC--GTHGVCIGGAC----- 713  
 Db 216 GPECQOQCEYGRCSAASGECTCPPGFRGARCELPCPAGSHGVQCAHSCKRKHNEPCS 275  
 Qy 714 -----RCEGWTAACDQ-----RVCHPRC--IEHG-TCK--DGKCE-CREGWNG 752  
 Db 276 PDTGSCSECEPGWNGTCCQCPCLPGTFGESECEQCPCHRGHEACEPDTGHCQRCDDPGWLG 335  
 Qy 753 EHC-----TIGRQTAGTETDGPDLGNGRCRTLQNSHQVCYQTCGRGPGCNVAMET 805  
 Db 336 PRCEDPCPTGTFEGDCGST-----CPTVOGSGDVTVG-----DCVCSAGYWGSPSCNACPA 387  
 Qy 806 SCADNKNE-----GDGLVDCLDPCCLQSAQNSLLRCGRSDPLDIIQOGOT----- 853  
 Db 388 GFHGNCSVPCEPEGLCHPVSGQPCGSGSRDTALIVSLVPLLLFLGLACCACCCWA 447  
 Qy 854 -----DMPAVK-SFYDRIKL-----LAGKDS-----HIIP-----GENPNS 885  
 Db 448 PRSDLKDRPADGATVSRMKLVQVWGTLTSLGSTLPCRSLSSHKLPWVTVSHHDPEVPFNH 507  
 Qy 886 SLVSLIRQVVTDTGTPLVGVNVPVKYKYG-----YTTTRDQGFDLIANGASLT 938  
 Db 508 SFIEPPSAGWATDD-----SFSSDPESGEADEVPAYCVPPQEGWVPAQAQSGSEAS 558  
 Qy 939 L 939  
 Db 559 L 559  
 RESULT 14  
 NOTC\_DROME  
 ID NOTC\_DROME STANDARD; PRT; 2703 AA.  
 AC P07207; P04154; O97458; Q9W4T8;  
 DT 01-NOV-1986 (Rel. 03, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Neurogenic locus Notch protein precursor.  
 GN N OR EG:140G11.1 OR EG:163A10.2 OR CG3936.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Oregon-R; TISSUE=Embryo;  
 RX MEDLINE=86079539; PubMed=3935325;  
 RA Wharton K.A., Johansen K.M., Xu T., Artavanis-Tsakonas S.;  
 RT "Nucleotide sequence from the neurogenic locus notch implies a gene  
 RT product that shares homology with proteins containing EGF-like  
 RT repeats."  
 RL Cell 43:567-581(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RC STRAIN=Canton-S, and Oregon-R; TISSUE=Embryo;  
RX MEDLINE=87064624; PubMed=3097517;  
RA Kidd S., Kelley M.R., Young M.W.;  
RT "Sequence of the notch locus of *Drosophila melanogaster*: relationship  
of the encoded protein to mammalian clotting and growth factors.";  
RL Mol. Cell. Biol. 5:3094-3108(1986).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX STRAIN=Berkeley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy K., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley J.S., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX STRAIN=Oregon-R;  
RX MEDLINE=20196011; PubMed=10731137;  
RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,  
RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,  
RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,  
RA Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,  
RA Pagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,  
RA Padogelli J., Peter A., Schoettler P., Werner M., Mourkioti F.,  
RA Beinert N., Dowe G., Schaefer U., Jaecckle H., Bucheton A.,  
RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,  
RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,  
RA Glover D.M.;  
RT "From sequence to chromosome: the tip of the X chromosome of *D. melanogaster*.";  
RL Science 287:2220-2222(2000).  
RN [5]  
RP SEQUENCE OF 2505-2611 FROM N.A.  
RX MEDLINE=85099329; PubMed=2981631;  
RA Wharton K.A., Redvick B., Finnerty V.G., Artavanis-Tsakonas S.;  
RT "opa: a novel family of transcribed repeats shared by the Notch locus  
and other developmentally regulated loci in *D. melanogaster*.";  
RL Cell 40:55-62(1985).  
RN [6]

RP SEQUENCE OF 1-8 FROM N.A.  
RX MEDLINE=87257846; PubMed=3037327;  
RA Kelley M.R., Kidd S., Berg R.L., Young M.W.;  
RT "Restriction of p-element insertions at the Notch locus of *Drosophila melanogaster*.";  
RL Mol. Cell. Biol. 7:1545-1548(1987).  
RN [7]  
RP REVIEW.  
RX Harris W.A.;  
RT "Many cell types specified by Notch function.;"  
RL Curr. Biol. 1:120-122(1991).  
CC -!- FUNCTION: NOTCH PROTEIN IS ESSENTIAL FOR PROPER DIFFERENTIATION OF  
ECTODERM.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- MISCELLANEOUS: SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO  
THE INNER PART OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS  
DEVELOPMENT IN INSECTS, THIS PROCESS IS UNDER CONTROL OF THE  
NEUROGENIC GENES.  
CC -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.  
CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.  
CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or  
send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M16152; AAB59220.1; -  
CC EMBL; M16153; AAB59220.1; JOINED.  
CC EMBL; M16149; AAB59220.1; JOINED.  
CC EMBL; M16150; AAB59220.1; JOINED.  
CC EMBL; M16151; AAB59220.1; JOINED.  
CC EMBL; K03508; AAB28725.1; -  
CC EMBL; M13889; AAB28725.1; JOINED.  
CC EMBL; K03507; AAB28725.1; JOINED.  
CC EMBL; AE003426; AAF45848.2; -  
CC EMBL; AL035436; CAB37610.1; -  
CC EMBL; AL035395; CAB37610.1; JOINED.  
CC EMBL; M12175; AAA74496.1; -  
CC EMBL; M16025; AAB28726.1; -  
CC PIR; A24420; A24420.  
CC PIR; A24768; A24768.  
CC PIR; A05267; A05267.  
CC HSP; P00740; IIXA.  
CC FlyBase; FBgn0004647; N.  
CC InterPro; IPR002110; ANK.  
CC InterPro; IPR000152; ASX\_hydroxyl.  
CC InterPro; IPR000561; EGF-like.  
CC InterPro; IPR000742; EGF 2.  
CC InterPro; IPR001881; EGF Ca.  
CC InterPro; IPR001438; EGF\_II.  
CC InterPro; IPR000800; Notch.  
CC Pfam; PF00023; ank; 6.  
CC Pfam; PF00008; EGF; 36.  
CC Pfam; PF00066; notch; 3.  
CC PRINTS; PR00010; EGFBL00D.  
CC PRINTS; PR01452; NOTCH.  
CC SMART; SM00248; ANK; 4.  
CC SMART; SM00179; EGF\_CA; 23.  
CC SMART; SM00001; EGF\_like; 13.  
CC SMART; SM00004; NL; 2.  
CC PROSITE; PS50088; ANK\_REPEAT; 5.  
CC PROSITE; PS0297; ANK\_REPEAT\_REGION; 1.  
CC PROSITE; PS00010; ASX\_HYDROXYL; 22.  
CC PROSITE; PS00022; EGF\_1; 34.  
CC PROSITE; PS01186; EGF\_2; 28.  
CC PROSITE; PS01187; EGF\_CA; 22.  
CC Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;  
KW Transmembrane; Signal; Glycoprotein.



FT SIGNAL 1 44 POTENTIAL.  
FT CHAIN 45 1703 NEUROGENIC LOCUS NOTCH PROTEIN.  
FT DOMAIN 45 1745 EXTRACELLULAR (POTENTIAL).  
FT TRANSEM 1746 1766 POTENTIAL.  
FT DOMAIN 1767 2703 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 58 95 EGF-LIKE 1.  
FT DOMAIN 96 136 EGF-LIKE 2.  
FT DOMAIN 139 176 EGF-LIKE 3.  
FT DOMAIN 177 215 EGF-LIKE 4.  
FT DOMAIN 217 253 EGF-LIKE 5. CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 255 291 EGF-LIKE 6.  
FT DOMAIN 293 329 EGF-LIKE 7. CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 331 370 EGF-LIKE 8. CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 372 408 EGF-LIKE 9. CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 409 447 EGF-LIKE 10.  
FT DOMAIN 449 486 EGF-LIKE 11. CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 488 524 EGF-LIKE 12. CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 526 562 EGF-LIKE 13. CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 564 600 EGF-LIKE 14. CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 602 637 EGF-LIKE 15. CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 639 675 EGF-LIKE 16. CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 677 713 EGF-LIKE 17. CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 715 751 EGF-LIKE 18. CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 753 789 EGF-LIKE 19. CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 791 827 EGF-LIKE 20. CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 829 865 EGF-LIKE 21. CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 867 905 EGF-LIKE 22.  
FT DOMAIN 907 944 EGF-LIKE 23. CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 946 982 EGF-LIKE 24.  
FT DOMAIN 984 1020 EGF-LIKE 25.  
FT DOMAIN 1022 1058 EGF-LIKE 26. CALCIUM-BINDING (POTENTIAL).

Query Match 2.8% Score 413; DB 1; Length 2703;  
Best Local Similarity 27.6% Pred. No. 3.7e-14;  
Matches 112; Conservative 45; Mismatches 143; Indels 106; Gaps 21;

Qy 531 SVQDCPRN-CHNGECVSV-CHCPGFLGADCAK-----RACPVLCGNGQYSGKT 580  
Db 602 NINDCSNPCH-RGKCIDVNSFRCLCDPGTYGTCOKQINECESNPQCQFDGHCQDRVGS 660  
Qy 581 --CCYSGKMGAECDVPMNOCIDPSCGGHSCIDG---NCVCSAGYKGEHCEE-VD-CL 632  
Db 661 YQCQAGTSKNGEVNVNCHSNPCNNGATCIDGINSYKQCQVPGFTGQHCERNVDCEI 720  
Qy 633 DPTCSSHGVC---VNGE-----CLCS 650  
721 SSPCANNGVCIDQVNGYKCECPRGFYDAHCLSDVDECASNPCVNEGRCEGDI NEFICHCP 780  
651 PGWGLNCELARVQCPDQCQSHGYLPD---TGLSCDPNWMGPDG--SVEVCSDV-CGT 704  
Db 781 PGYTKRCLELDICSSNPCHGGTCTYDKLNAFSCQCMPGYTGQKCEINDDCVTNPCGN 840  
Qy 705 HGVCI---GGACRCEEGWTGAACDQV---CHPRCIEHGTC-----DGKCECREGWN 751  
Db 841 GGTCLDKVNGYKCVKVPFTGRDESKMDPCASNRCKNEAKCTPSSNFDSCCTCKLGYT 900  
Qy 752 GEHCTIGROTAGTDCG--PDLGNGRCRTLGONSQWCVCOTGWRGPGCNVAMETSCAD 809  
Db 901 GRYC-----DEDIDECSLSPCRNGASCLNVPGRYCRCTKGTGEGRDCAINTDDCASF 953  
Qy 810 NKDNEG---DGLVD--CL-----DPDCCLQSAACONSLLC 838  
Db 954 PCQNGGTCLDGDGYSLCVDFDGGKHCETDINECLSPCQNGATC 999

RESULT 15

GLP1\_CAEEL STANDARD; PRT; 1295 AA.  
ID GLP1\_CAEEL  
AC P13508;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Glp-1 protein precursor.

GN GLP-1 OR EMB-33 OR F02A9.6.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=89336787; PubMed=2758466;  
RA Yochem J., Greenwald I.;  
RT "glp-1 and lin-12, genes implicated in distinct cell-cell  
RL interactions in C. elegans, encode similar transmembrane proteins.";  
RN Cell 58:553-563(1989).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,  
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,  
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
RA Sulston J., Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K.,  
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
RA Wohldman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RL elegans.";  
RN Nature 368:32-38(1994).  
[3]  
RP DELETION OF 1174-1295.  
RX MEDLINE=91351288; PubMed=1881436;  
RA Mango S.E., Maine E.M., Kimble J.;  
RT "Carboxy-terminal truncation activates glp-1 protein to specify  
RL vulval fates in Caenorhabditis elegans.";  
RN Nature 352:811-815(1991).  
[4]  
RP CHARACTERIZATION OF FUNCTION OF THE ANK-REPEATS.  
RX MEDLINE=93354444; PubMed=8350921;  
RA Roehl H., Kimble J.;  
RT "Control of cell fate in C. elegans by a GLP-1 peptide consisting  
RL primarily of ankyrin repeats.";  
RN Nature 364:632-635(1993).  
[5]  
RP FUNCTION.  
RX MEDLINE=94208066; PubMed=8156602;  
RA Mello C.C., Draper B.W., Priess J.R.;  
RT "The maternal genes apx-1 and glp-1 and establishment of  
RL dorsal-ventral polarity in the early C. elegans embryo.";  
RN Cell 77:95-106(1994).  
CC -1- FUNCTION: INVOLVED IN THE SPECIFICATION OF THE CELL FATES OF THE  
CC BLASTOMERES ABA AND APA. PROPER SIGNALING BY GLP-1 INDUCES ABA  
CC DESCENDANTS TO PRODUCE ANTERIOR PHARYNGEAL CELLS, AND APA  
CC DESCENDANTS TO ADOPT A DIFFERENT FATE. CONTRIBUTES TO THE  
CC ESTABLISHMENT THE DORSAL-VENTRAL AXIS IN EARLY EMBRYOS.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- DEVELOPMENTAL STAGE: ACTS ON ABA DEVELOPMENT DURING 4-CELL AND  
CC 12-CELL STAGES, AND ON ABA DEVELOPMENT DURING 12-CELL AND 28-CELL  
CC STAGES.  
CC -1- SIMILARITY: HIGH, TO C.ELEGANS LIN-12.  
CC -1- SIMILARITY: CONTAINS 10 EGF-LIKE DOMAINS.  
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.  
CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M25080; AAA28058.1; -  
DR EMBL; Z19555; CAA79620.1; -  
DR EMBL; Z29116; CAA79620.1; JOINED.  
DR EMBL; Z29116; CAA82373.1; -  
DR EMBL; Z19555; CAA82373.1; JOINED.  
DR PIR; A32901; A32901.  
DR HSP; P00740; 1EDM.  
DR WormPep; F02A9.6; CE00237.  
DR InterPro; IPR002110; ANK.  
DR InterPro; IPR000152; Asx\_hydroxyl.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR000742; EGF\_2.  
DR InterPro; IPR001881; EGF\_Ca.  
DR InterPro; IPR000800; Notch.  
DR Pfam; PF00023; ank; 6.  
DR Pfam; PF00008; EGF; 10.  
DR Pfam; PF00066; notch; 3.  
DR PRINTS; PR01452; NOTCH.  
DR SMART; SM00248; ANK; 4.  
DR SMART; SM00179; EGF\_Ca; 1.  
DR SMART; SM00001; EGF\_Like; 7.  
DR SMART; SM00004; NL; 3.  
DR PROSITE; PS50088; ANK\_REPEAT; 3.  
DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 2.  
DR PROSITE; PS00022; EGF\_1; 10.  
DR PROSITE; PS01186; EGF\_2; 8.  
DR PROSITE; PS01187; EGF\_Ca; 1.  
KW Differentiation; Repeat; ANK repeat; EGF-like domain; Transmembrane;  
KW Glycoprotein; Signal.  
FT SIGNAL 1 15  
FT CHAIN 16 1295  
FT DOMAIN 16 764  
FT TRANSMEM 765 786  
FT DOMAIN 787 1295  
FT DOMAIN 19 58  
FT DOMAIN 117 152  
FT DOMAIN 154 190  
FT DOMAIN 190 230  
FT DOMAIN 232 269  
FT DOMAIN 271 308  
FT DOMAIN 316 359  
FT DOMAIN 369 406  
FT DOMAIN 407 443  
FT DOMAIN 446 479  
FT REPEAT 493 527  
FT REPEAT 528 568  
FT REPEAT 569 608  
FT REPEAT 961 990  
FT REPEAT 994 1023  
FT REPEAT 1030 1062  
FT REPEAT 1074 1103  
FT REPEAT 1107 1136  
FT DISULFID 23 35  
FT DISULFID 29 46  
FT DISULFID 48 57  
FT DISULFID 121 131  
FT DISULFID 126 140  
FT DISULFID 142 151  
FT DISULFID 158 169  
FT DISULFID 163 178  
FT DISULFID 180 189  
FT DISULFID 194 206  
FT DISULFID 201 218  
FT DISULFID 220 229  
FT DISULFID 236 248  
FT DISULFID 242 257  
FT DISULFID 259 268  
FT DISULFID 275 286  
FT DISULFID 280 296  
FT DISULFID 298 307  
FT DISULFID 373 384

FT DISULFID 378 394 BY SIMILARITY.  
FT DISULFID 396 405 BY SIMILARITY.  
FT DISULFID 411 422 BY SIMILARITY.  
FT DISULFID 416 431 BY SIMILARITY.  
FT DISULFID 433 442 BY SIMILARITY.  
FT DISULFID 450 461 BY SIMILARITY.  
FT DISULFID 455 467 BY SIMILARITY.  
FT DISULFID 469 478 BY SIMILARITY.  
FT CARBOHYD 244 244 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 245 245 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 333 333 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 381 381 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 609 609 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 675 675 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 1295 AA; 422AAD0A2DEEF3B4 CRC64;  
  
Query Match 2.7%; Score 399; DB 1; Length 1295;  
Best Local Similarity 20.3%; Pred.No. 6.7e-14;  
Matches 233; Conservative 139; Mismatches 370; Indels 408; Gaps 57;  
  
QY 530 DSVDQCPRN-CHGNBCEVSGV-----CHCFPGFLGADCAKAAACPVL-----CSGNGOY 576  
DB 153 EGDHCAQNECAEGSTCVSNVYNYCDPIGKSGRYCERTECALMGNICNHGRCIPNRE 212  
QY 577 SKG-TQCYSGWMKGAECDDVPMQC-IDPCGGHSGCI-----DGNVCVSAGYKGEHCEE-V 629  
DB 213 DKNFRVCDSGYEGEFCNKNKECLTEETCVNNSTCFNLHGDFTCTCKPGYAGKYCEEAI 272  
QY 630 D-CLDPTCSSHGVNCE-----CLCSPWGGGLNCELARVQCPD-----QC 669  
DB 273 DMCKDYVCNDGYCAHDSNQMFICYCEQGTQORCE---IECPGSGGIHCDLPLQRPIC 329  
QY 670 S-GHGTYLPD---TGLCSCDPNMWMPDC-----SVEVCSVD-CGTHGVCI----- 709  
DB 330 SRNGTCYNDGRINGFCVCEPDYIGDRCINRKPDKFPDIOSCKYPCVNNATCIDLN 389  
QY 710 -GGACRCBEGWTGAACDQR-VCHP-RCIEHGTC--DG--KCECRGWNGEHCTI--GROT 761  
DB 390 SGYSCHCPLGFGYGLNCEQHLCTPTTCANGTCEGVNGVIRNCNPGFSGDYCEIKDROL 449  
QY 762 AGTETDGCPLDNGNGRCITLGNSSQCVCTGWRGPGCN--VAMETS-----CADNK 811  
DB 450 CSRHP-----CKNGVC---ANTGYCECYGTGTCEEVLVIEKSKETVINDLCEQRK 500  
QY 812 --DNEGDLVDCIDPDCCIQSACQNSLLRCGRSDPLDIIQOGQTDMPA----- 857  
DB 501 CMDLASNGI---CNPECNLEECNFDGDCSGGQRPFSKCQ-----YPARCAQFANGVCN 552  
QY 858 -----VKSFYDRIKLAGKDSHIIPGENPFFNSLV---SLIRGVVTTDGTPLVGVNVS 909  
DB 553 QECNNECLYD-----GLDC-----QSELPFPAHIRKHKCIERRGDGVNCLCS 596  
QY 910 FVKYPKYG-----YTITRODGTPLIANGGASL-----TLHEFR--- 943  
DB 597 FICGDFDGGDCNNGTEAIIISDIRIKVQIDPIEFQATGGETLMQISANLRATVRIORDEL 656  
QY 944 APFMSQERTVWLPWNSFYAMDITLVKMTENSIIPSCDLSGFVFRPDPDIIISPLSTFFSAAP 1003  
DB 657 GPLVFR-----WDGEHEMERVEMNSSK-----LEDQFVLSHHVRRYRQA-- 695  
QY 1004 GONPIVETQVHLHEEIELPGSNVKLYLSRPTAGYKSLKTKMTQSTVPLNLRVILMVA 1063  
DB 696 -----VVTGLVLYLEVE---EICKPEFCRFSTA-----QSVVDL----- 726  
QY 1064 VEGHLFQKSFQASPNLASTFIWDKTDAYGORVYGLSDAVSVGVFEYETCPSLILWEKRTA 1123  
DB 727 -----IAAGLV-----KSDGRMSUG-----LPITAMVA 750  
QY 1124 LLQGFELDPNLSGGWSLKHHL-----NVKSGILHKGTE-----NOFLTOQ 1166  
DB 751 VPKRNEIDE-----GWSRSQVILFACIAFLAFGVVAGVIKNGPERSKRKRKMYNATVWMP 806

QY 1167 PAITSIMGRRRSISPCSCNGL-----AEGNKLLAPVALAV 1204  
Db 807 PHESTNEGR-RNQSNHSSQCSLLDNSAYHPNTRKHCSDYSTGYNGEQYSQIYPQTLAN 865  
QY 1205 GIDGSLYVDFNYIRIPPSRNVTSILELRNKEFKHSNPAHKYYLAVDPVSGSLYVSDT 1264  
Db 866 G-----YPGDYNELN-----FDFQSETFAPADLPADLPLHVQAAGPDAITAPI 909  
QY 1265 NSRRIYRKSLSGTKDLAGNSEWAGTGEQCLPFDEARCGDGGKAIDATLMSPRGIAVDK 1324  
Db 910 TNESYNQVDSKYRRRVLHHLAANVRGKPEDVITTEAIRCLKAGADYNA----- 957  
QY 1325 NGLMYFVDATMIRKVDQNGIISTLGSNDLTAVRPLSCDSSMDVAQVRELPDLDLVNPM 1384  
Db 958 -----RDCDENTAL-----MLAVR-----AHRVRL----- 977  
QY 1385 DNSLYVLENNVILRITENHOVSIIAGRPMHCQVPGIDYSLSKLATHSALESASATAISHT 1444  
978 -----SVLLREGAN-----PTIFNNSERSALHEAVVVKDLRLRH- 1013  
QY 1445 GVLYYTETDEKKINRLQVTTNG-----EICLLAGAAASDCDCKNDV 1485  
Db 1014 -----LLTDKRLKKEIDELDRNGMTALMLVARELKGHVEMAEALLSKGAKLDYDGAARK 1068  
QY 1486 NCNCYSGDDA 1495  
Db 1069 DSNKYKGRTA 1078

Search completed: September 18, 2002, 11:17:02  
Job time: 517 sec

